A Comprehensive Review of Modern Object Segmentation Approaches

Other titles in Foundations and Trends® in Computer Graphics and Vision

Deep Learning for Multimedia Forensics
Irene Amerini, Aris Anagnostopoulos, Luca Maiano and Lorenzo Ricciardi Celsi
ISBN: 978-1-68083-854-1

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art
Joel Janai, Fatma Güney, Aseem Behl and Andreas Geiger
ISBN: 978-1-68083-688-2

Discrete Graphical Models - An Optimization Perspective
Bogdan Savchynskyy
ISBN: 978-1-68083-638-7

Line Drawings from 3D Models: A Tutorial
Pierre Bénard and Aaron Hertzmann
ISBN: 978-1-68083-590-8

Publishing and Consuming 3D Content on the Web: A Survey
Marco Potenziani, Marco Callieri, Matteo Dellepiane and Roberto Scopigno
ISBN: 978-1-68083-536-6

Crowdsourcing in Computer Vision
Adriana Kovashka, Olga Russakovsky, Li Fei-Fei and Kristen Grauman
ISBN: 978-1-68083-212-9

A Comprehensive Review of Modern Object Segmentation Approaches

Yuanbo Wang
Invitae Corporation
mcdy143@gmail.com

Unaiza Ahsan
The Home Depot
unaiza_ahsan@homedepot.com

Hanyan Li
Indeed, Inc.
f1annlee@gmail.com

Matthew Hagen
Amazon.com, Inc.
mathage@amazon.com

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

Y. Wang *et al.*. *A Comprehensive Review of Modern Object Segmentation Approaches.* Foundations and Trends® in Computer Graphics and Vision, vol. 13, no. 2-3, pp. 111–283, 2022.

ISBN: 978-1-63828-071-2

Foundations and Trends® in Computer Graphics and Vision
Volume 13, Issue 2-3, 2022

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2022, Volume 13, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Contents

A Comprehensive Review of Modern Object Segmentation Approaches

Yuanbo Wang[1], Unaiza Ahsan[2], Hanyan Li[3] and Matthew Hagen[4]

[1]*Invitae Corporation, USA; mcdy143@gmail.com*
[2]*The Home Depot, USA; unaiza_ahsan@homedepot.com*
[3]*Indeed Inc., USA; f1annlee@gmail.com*
[4]*Amazon.com, Inc., USA; mathage@amazon.com*

ABSTRACT

Image segmentation is the task of associating pixels in an image with their respective object class labels. It has a wide range of applications in many industries including healthcare, transportation, robotics, fashion, home improvement, and tourism. Many deep learning-based approaches have been developed for image-level object recognition and pixel-level scene understanding — with the latter requiring a much denser annotation of scenes with a large set of objects. Extensions of image segmentation tasks include 3D and video segmentation, where units of voxels, point clouds, and video frames are classified into different objects. We use "Object Segmentation" to refer to the union of these segmentation tasks. In this monograph, we investigate both traditional and modern object segmentation approaches, comparing their strengths, weaknesses, and utilities. We examine in detail the wide range of deep learning-based segmentation techniques developed in recent years, provide a review of

Yuanbo Wang, Unaiza Ahsan, Hanyan Li and Matthew Hagen (2022), "A Comprehensive Review of Modern Object Segmentation Approaches", Foundations and Trends® in Computer Graphics and Vision: Vol. 13, No. 2-3, pp 111–283. DOI: 10.1561/0600000097.

the widely used datasets and evaluation metrics, and discuss potential future research directions.

1

Introduction

1.1 Overview

Automated visual recognition tasks such as image classification, image captioning, object detection and image segmentation are essential for image and video processing. Before the advent of deep neural networks, traditional techniques leveraged hand-crafted heuristics to extract visual features and manually tuned parameters to combine these features for inferences and decisions. These techniques are simple yet effective for many cases. However, they are often not generalizable and difficult to configure. For example, Canny Edge Detection, an algorithm that uses Gaussian filters and thresholding to identify edges in images have two key adjustable parameters-size of filters and edge strength thresholds that need to be tuned. If not selected carefully, these parameters can greatly impact the effectiveness of the algorithm, resulting in either missing or false positive edges.

1.2 Convolutional Neural Networks

With the advent of the deep learning era, more powerful and accurate automated visual recognition techniques were enabled by deep neural

networks that can handle high-dimensional data. In particular, convolutional neural networks (CNN) (LeCun *et al.*, 1999) excel at image recognition and pattern identification due to its ability to capture space invariance of shapes in images. CNNs achieve this through two major components: the convolutional layer, pooling layer, and non-linear layers. The convolutional layers apply sliding kernels to extract features from the input image, starting from low-level features such as edges in the earlier layers, to high-level features such as eyes and nose in the later layers. The pooling layer then reduces the resolution of the output feature maps from the convolutional layers, allowing the network to achieve translational and deformation invariance. LeNet-5, developed by LeCun *et al.* (1998), is the first convolutional neural network applied to the famous handwritten digit recognition task MNIST.[1] Although only consisting of a pair of consecutive convolutional and pooling layers followed by three fully connected layers, LeNet-5 is able to achieve a 98.49% accuracy on the MNIST dataset. However, when applied to The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Deng *et al.*, 2009) task, the test accuracy drops to only 66%. Compared with the human level accuracy of 94%, LeNet-5 is not sufficient for more challenging visual recognition tasks.

In 2012, a breakthrough was made when Hinton's team won the ILSVRC challenge with deep learning (Krizhevsky *et al.*, 2012). The challenge included the largest image dataset of the time, totaling over one million images spanning 1,000 object categories. For the first three years since the challenge commenced, most visual recognition systems did not make any breakthrough at this image classification task. The deep learning model proposed by Krizhevsky *et al.* (2012) was named AlexNet, which is often considered the first deep convolutional network. It is similar in structure with LeNet-5, but consists of 3 more convolutional layers and has a total of 62 million trainable variables. It significantly boosted the performance of previous state-of-art image classification techniques by achieving considerably better classification accuracy in the ILSVRC challenge than the second place, reducing the top-5 error rate by over 10% (Krizhevsky *et al.*, 2012). It was not until then that

[1] http://yann.lecun.com/exdb/mnist/

deep learning started to take over as the go-to approach for challenging computer vision tasks.

Despite the significant improvement AlexNet made towards image classification, it uses too many parameters and is therefore intractable for large-scale training. VGGNet (Simonyan and Zisserman, 2014) solves this problem by replacing the large kernel-sized convolutional layers with multiple layers containing 3 x 3 kernels. The network contains a total of 13 convolutional layers and 3 fully connected layers. The authors demonstrate that many small-sized filters achieve the same functions as fewer filters with large kernel sizes, but with much fewer parameters.

As CNNs continue to develop and grow deeper, a problem that often arises during the training stage is vanishing gradient – weights disappearing during back-propagation, which hinders the performance of models. ResNet (He *et al.*, 2016), developed one year after VGGNet (Simonyan and Zisserman, 2014), leverages two types of skip connections – identity and projection, to tackle this issue. In addition, it also uses batch normalization (Ioffe and Szegedy, 2015) to stabilize and increase training speed.

Developed by Google, The Inception (Szegedy *et al.*, 2015) network is aimed to capture sparse salient features of varying sizes. It achieves this through stacking multiple sizes (5 x 5, 3 x 3, and 1 x 1) of kernels at the same layer and combining outputs to feed to the next layer. 1 x 1 convolutions are used to reduce the number of channels, thereby reducing the number of parameters of the network.

1.3 Object Detection

A precursor to the image segmentation problem, object detection not only classifies, but also localizes each object in an image. One of the earlier approaches to solving object detection is the sliding window approach (Bosch *et al.*, 2007; Chum and Zisserman, 2007; Dalal and Triggs, 2005; Ferrari *et al.*, 2007; Rowley *et al.*, 1995). By taking different crops from an input image and applying CNNs to classify cropped regions (a cropped region is classified as "background" if no target object is identified), different objects are simultaneously localized and categorized from the input image.

However, because objects appear in various sizes and in many locations in the image, it is very computationally expensive to apply CNNs to a potentially infinite set of cropped input regions. Instead, a more scalable approach called region proposals (Alexe *et al.*, 2012; Uijlings *et al.*, 2013; Zitnick and Dollár, 2014) was developed to find candidate regions in images that are likely to contain objects quickly based on traditional image and signal processing techniques. One example of such a technique is Selective Search, which uses sub-segmentation and greedy search to generate and recursively combine regions into a final list of 2,000 candidate regions (Uijlings *et al.*, 2013). With region proposals approaches, CNNs can then be applied to a much smaller set of cropped regions to produce final detection outputs, thus making this task much more computationally tractable.

The idea of combining region proposals and CNNs for object detection first materialized in Girshick *et al.* (2014). After the features are extracted from the proposed regions using CNNs, an SVM is used to 1) determine whether an object is present in the proposed region and classifies it if present, and 2) perform bounding box regression to compute offsets to the proposed region and refine its boundary. Although R-CNN is able to resolve the challenges with the sliding window approach to object detection, it is still very slow since CNNs need to be applied to 2,000 regions, and selective search could generate bad proposals since it is a fixed algorithm.

The same authors of R-CNN developed Fast R-CNN (Girshick, 2015) to address the slow selective search stage. Instead of applying CNNs to proposed regions, the input image is fed into the CNNs directly to generate a feature map. From the feature map, candidate regions are produced and an RoI pooling layer is then applied to reshape these regions into fixed sizes before they are fed into a fully connected region to produce a feature vector. Finally, a softmax layer is applied to classify this feature vector into objects, and a bounding box regressor is used to compute the offset values to refine boundaries. This is a much faster approach than R-CNN because the CNNs need only be run one time on the input image instead of on 2,000 regions. Although Fast-R-CNN is significantly faster than R-CNN during both training and testing, the region proposal stage remains to be the bottleneck. This is because it

still uses selective search on the CNN feature map to propose candidate regions.

Faster R-CNN (Ren *et al.*, 2015) is a similar approach to both algorithms above, but leverages a separate network on the CNN feature map to propose candidate regions instead of using selective search. As a result, Faster R-CNN beats its predecessors in inference speed and can generate detection results in real-time.

Whereas the object detection models from the R-CNN family use regions to identify objects in an image, YOLO (Redmon *et al.*, 2016) is an approach that looks at not just the regions of interest, but the entire image. The image is divided into grids. A set of base boxes centered at each grid cell is fed as input to a single CNN to predict both object class scores and bounding boxes. This approach is extremely fast as detection is reframed as a regression problem. YOLO also sees the entire image when making predictions and makes fewer mistakes compared to Fast R-CNN when identifying background patches (Redmon *et al.*, 2016). However, it is not good at predicting small objects due to the constraints placed on the base boxes.

Similar to YOLO, SSD (Liu *et al.*, 2016a) is a single-shot (achieves object localization and detection simultaneously) object detector with fast inference speed to enable real-time detection. It uses a modified version of MultiBox (Erhan *et al.*, 2014), a bounding box regression technique, to generate candidate bounding boxes on each grid cell of the feature map. Given a set of default bounding boxes with different aspect ratios, SSD selects those with an intersection over union (IOU) score larger than 0.5 with the ground truth bounding box as candidates and performs bounding box regression to refine the boundaries. It also predicts, for each object class c and each candidate bounding box b, the probabilities of box b containing object c.

1.4 Object Segmentation

While object detection aims only to localize objects with bounding boxes and classifying localized objects, object segmentation adds the task of delineating precise object boundaries by classifying every unit in an input to an object class or the background. The general formulation for this

problem can be defined as follows: given an input space $\mathcal{X}$, and a label space $\mathcal{L}$, object segmentation finds a mapping f: $\mathcal{X} \mapsto \mathcal{L}$. It has a wide range of applications in many industries including healthcare (e.g. cancer detection), transportation (self-driving), robotics (guidance), fashion (virtual try-ons), home improvement (decor visualizers), and tourism (destination recognition). Segmentation of an image breaks it down into simplified representations that can aid downstream interpretation tasks. Thus, segmentation differs from detection as it does not have to be tied to a specific task, but rather an intermediate step in machine perception. Nevertheless, it is more challenging and usually more time-consuming than object detection, thereby requiring more advanced techniques and more high-quality annotated training data. Object segmentation is fundamentally tied to human image perception as recognition of objects from an input image can be conceptualized into a process of dissecting regions of deep concavity into simple volumetric components such as blocks, wedges, and cones (Biederman, 1985).

Recognition-by-components (RBC) theory states that robust object perception is possible because detection of edge properties such as curvature, symmetry, and parallelism is usually invariant over image quality and viewing angle (Biederman, 1985). However, because machines process images through various means unlike the process of human interpretation, it is far from enough to only focus on these components when performing object segmentation. In this monograph, we review various techniques developed over the years for the task of object segmentation, with a focus on the deep learning-based techniques for the two most widely solved segmentation tasks: Semantic Segmentation and Instance Segmentation. We also survey methods developed for Panoptic, Video and 3D data segmentation. We categorize and compare various techniques developed for each task. The readers will find that various themes emerge from these techniques that push machines to their limits and often times deviate from human perception principles. In addition, we provide an overview of the widely used benchmark datasets for each of these techniques, along with the respective evaluation metrics to measure the models' performances. Finally, we discuss potential directions for future research in these areas.

References

Abhishek, K. and G. Hamarneh. (2019). "Mask2lesion: Mask-constrained adversarial skin lesion image synthesis". In: *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer. 71–80.

Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. (2010). "Slic superpixels". *Tech. rep.*

Adams, R. and L. Bischof. (1994). "Seeded region growing". *IEEE Transactions on pattern analysis and machine intelligence*. 16(6): 641–647.

Akbarimoghaddam, P., A. Ziaei, and H. Azarnoush. (2022). "Deep active contours using locally controlled distance vector flow". *Signal, Image and Video Processing*. DOI: 10.1007/s11760-022-02134-1.

Akhtar, N. and A. Mian. (2018). "Threat of adversarial attacks on deep learning in computer vision: A survey". *Ieee Access*. 6: 14410–14430.

Alexe, B., T. Deselaers, and V. Ferrari. (2012). "Measuring the objectness of image windows". *IEEE transactions on pattern analysis and machine intelligence*. 34(11): 2189–2202.

Alvarez, J. M., T. Gevers, Y. LeCun, and A. M. Lopez. (2012). "Road scene segmentation from a single image". In: *European Conference on Computer Vision*. Springer. 376–389.

Arbelaez, P., M. Maire, C. Fowlkes, and J. Malik. (2010). "Contour detection and hierarchical image segmentation". *IEEE transactions on pattern analysis and machine intelligence*. 33(5): 898–916.

Arbeláez, P., J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik. (2014). "Multiscale combinatorial grouping". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 328–335.

Armeni, I., S. Sax, A. R. Zamir, and S. Savarese. (2017). "Joint 2d-3d-semantic data for indoor scene understanding". *arXiv preprint arXiv:1702.01105*.

Armeni, I., O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. (2016). "3d semantic parsing of large-scale indoor spaces". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1534–1543.

Badrinarayanan, V., A. Kendall, and R. Cipolla. (2017). "Segnet: A deep convolutional encoder-decoder architecture for image segmentation". *IEEE transactions on pattern analysis and machine intelligence*. 39(12): 2481–2495.

Bai, M. and R. Urtasun. (2017). "Deep watershed transform for instance segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5221–5229.

Behley, J., M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. (2019). "Semantickitti: A dataset for semantic scene understanding of lidar sequences". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9297–9307.

Bell, S., P. Upchurch, N. Snavely, and K. Bala. (2015). "Material recognition in the wild with the materials in context database". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3479–3487.

Bian, X., S. N. Lim, and N. Zhou. (2016). "Multiscale fully convolutional network with application to industrial inspection". In: *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 1–8.

Biederman, I. (1985). "Human image understanding: Recent research and a theory". *Computer vision, graphics, and image processing*. 32(1): 29–73.

Blanchon, M., O. Morel, Y. Zhang, R. Seulin, N. Crombez, and D. Sidibé. (2019). "Outdoor Scenes Pixel-wise Semantic Segmentation using Polarimetry and Fully Convolutional Network." In: *VISIGRAPP (5: VISAPP)*. 328–335.

Blin, R., S. Ainouz, S. Canu, and F. Meriaudeau. (2019). "Road scenes analysis in adverse weather conditions by polarization-encoded images and adapted deep learning". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE. 27–32.

Blin, R., S. Ainouz, S. Canu, and F. Meriaudeau. (2020). "A new multimodal RGB and polarimetric image dataset for road scenes analysis". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 216–217.

Blott, G., M. Takami, and C. Heipke. (2018). "Semantic segmentation of fisheye images". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 0–0.

Bolya, D., C. Zhou, F. Xiao, and Y. J. Lee. (2019). "Yolact: Real-time instance segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9157–9166.

Bolya, D., C. Zhou, F. Xiao, and Y. J. Lee. (2020). "Yolact++: Better real-time instance segmentation". *IEEE transactions on pattern analysis and machine intelligence*.

Bosch, A., A. Zisserman, and X. Munoz. (2007). "Representing shape with a spatial pyramid kernel". In: *Proceedings of the 6th ACM international conference on Image and video retrieval*. 401–408.

Brodeur, S., E. Perez, A. Anand, F. Golemo, L. Celotti, F. Strub, J. Rouat, H. Larochelle, and A. Courville. (2017). "Home: A household multimodal environment". *arXiv preprint arXiv:1711.11017*.

Brostow, G. J., J. Fauqueur, and R. Cipolla. (2009). "Semantic object classes in video: A high-definition ground truth database". *Pattern Recognition Letters*. 30(2): 88–97.

Byeon, W., T. M. Breuel, F. Raue, and M. Liwicki. (2015). "Scene labeling with lstm recurrent neural networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3547–3555.

Caesar, H., V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. (2020). "nuscenes: A multimodal dataset for autonomous driving". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11621–11631.

Caesar, H., J. Uijlings, and V. Ferrari. (2018). "Coco-stuff: Thing and stuff classes in context". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1209–1218.

Cai, Z. and N. Vasconcelos. (2019). "Cascade r-cnn: High quality object detection and instance segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Canny, J. (1986). "A computational approach to edge detection". *IEEE Transactions on pattern analysis and machine intelligence*. (6): 679–698.

Cao, J., R. M. Anwer, H. Cholakkal, F. S. Khan, Y. Pang, and L. Shao. (2020). "Sipmask: Spatial information preservation for fast image and video instance segmentation". In: *European Conference on Computer Vision*. Springer. 1–18.

Cao, J., H. Leng, D. Lischinski, D. Cohen-Or, C. Tu, and Y. Li. (2021). "ShapeConv: Shape-aware Convolutional Layer for Indoor RGB-D Semantic Segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7088–7097.

Carion, N., F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. (2020). "End-to-end object detection with transformers". In: *European Conference on Computer Vision*. Springer. 213–229.

Chan, T. F. and L. A. Vese. (2001). "Active contours without edges". *IEEE Transactions on image processing*. 10(2): 266–277.

Chandra, S., C. Couprie, and I. Kokkinos. (2018). "Deep spatio-temporal random fields for efficient video segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8915–8924.

Chang, A. X., T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, *et al.* (2015). "Shapenet: An information-rich 3d model repository". *arXiv preprint arXiv:1512.03012*.

Chang, C.-Y., S.-E. Chang, P.-Y. Hsiao, and L.-C. Fu. (2020). "EPSNet: Efficient Panoptic Segmentation Network with Cross-layer Attention Fusion". In: *Proceedings of the Asian Conference on Computer Vision (ACCV)*.

Changhee, H., M. Kohei, S. Shin'ichi, and N. Hideki. (2019). "Learning more with less: GAN-based medical image augmentation". *Med. Imaging Technol.* 6.

Chartsias, A., T. Joyce, R. Dharmakumar, and S. A. Tsaftaris. (2017). "Adversarial image synthesis for unpaired multi-modal cardiac data". In: *International workshop on simulation and synthesis in medical imaging*. Springer. 3–13.

Chen, H., C. Shen, and Z. Tian. (2020a). "Unifying Instance and Panoptic Segmentation with Dynamic Rank-1 Convolutions". *arXiv preprint arXiv:2011.09796*.

Chen, H., K. Sun, Z. Tian, C. Shen, Y. Huang, and Y. Yan. (2020b). "BlendMask: Top-down meets bottom-up for instance segmentation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8573–8581.

Chen, K., J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang, *et al.* (2019a). "Hybrid task cascade for instance segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4974–4983.

Chen, L.-C., J. T. Barron, G. Papandreou, K. Murphy, and A. L. Yuille. (2016a). "Semantic image segmentation with task-specific edge detection using cnns and a discriminatively trained domain transform". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4545–4554.

Chen, L.-C., M. Collins, Y. Zhu, G. Papandreou, B. Zoph, F. Schroff, H. Adam, and J. Shlens. (2018a). "Searching for efficient multi-scale architectures for dense image prediction". *Advances in neural information processing systems*. 31.

Chen, L.-C., A. Hermans, G. Papandreou, F. Schroff, P. Wang, and H. Adam. (2018b). "Masklab: Instance segmentation by refining object detection with semantic and direction features". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4013–4022.

Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. (2017a). "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". *IEEE transactions on pattern analysis and machine intelligence*. 40(4): 834–848.

Chen, L.-C., G. Papandreou, F. Schroff, and H. Adam. (2017b). "Rethinking atrous convolution for semantic image segmentation". *arXiv preprint arXiv:1706.05587*.

Chen, L.-C., H. Wang, and S. Qiao. (2020c). "Scaling Wide Residual Networks for Panoptic Segmentation". *arXiv preprint arXiv:2011.11675*.

Chen, L.-C., Y. Yang, J. Wang, W. Xu, and A. L. Yuille. (2016b). "Attention to scale: Scale-aware semantic image segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3640–3649.

Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. (2018c). "Encoder-decoder with atrous separable convolution for semantic image segmentation". In: *Proceedings of the European conference on computer vision (ECCV)*. 801–818.

Chen, L.-Z., Z. Lin, Z. Wang, Y.-L. Yang, and M.-M. Cheng. (2021). "Spatial information guided convolution for real-time RGBD semantic segmentation". *IEEE Transactions on Image Processing*. 30: 2313–2324.

Chen, P., Q. Xiao, J. Xu, X. Dong, L. Sun, W. Li, X. Ning, G. Wang, and Z. Chen. (2020d). "Harnessing semantic segmentation masks for accurate facial attribute editing". *Concurrency and Computation: practice and experience*: e5798.

Chen, Q., A. Cheng, X. He, P. Wang, and J. Cheng. (2020e). "Spatialflow: Bridging all tasks for panoptic segmentation". *IEEE Transactions on Circuits and Systems for Video Technology*. 31(6): 2288–2300.

Chen, X., J. Wang, and M. Hebert. (2020f). "PanoNet: Real-time Panoptic Segmentation through Position-Sensitive Feature Embedding". *arXiv preprint arXiv:2008.00192*.

Chen, X., R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille. (2014). "Detect what you can: Detecting and representing objects using holistic models and body parts". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1971–1978.

Chen, X., K.-Y. Lin, J. Wang, W. Wu, C. Qian, H. Li, and G. Zeng. (2020g). "Bi-directional cross-modality feature propagation with separation-and-aggregation gate for RGB-D semantic segmentation". In: *European Conference on Computer Vision*. Springer. 561–577.

Chen, X., R. Girshick, K. He, and P. Dollár. (2019b). "Tensormask: A foundation for dense object segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2061–2069.

Chen, X., B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng. (2019c). "Learning active contour models for medical image segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11632–11640.

Cheng, B., M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L.-C. Chen. (2020). "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12475–12485.

Cheng, B., I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar. (2022a). "Masked-attention mask transformer for universal image segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1290–1299.

Cheng, B., A. Schwing, and A. Kirillov. (2021). "Per-pixel classification is not all you need for semantic segmentation". *Advances in Neural Information Processing Systems*. 34: 17864–17875.

Cheng, D., R. Liao, S. Fidler, and R. Urtasun. (2019). "Darnet: Deep active ray network for building segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7431–7439.

Cheng, J., Y.-H. Tsai, S. Wang, and M.-H. Yang. (2017a). "Segflow: Joint learning for video object segmentation and optical flow". In: *Proceedings of the IEEE international conference on computer vision*. 686–695.

Cheng, T., X. Wang, S. Chen, W. Zhang, Q. Zhang, C. Huang, Z. Zhang, and W. Liu. (2022b). "Sparse Instance Activation for Real-Time Instance Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4433–4442.

Cheng, Y., R. Cai, Z. Li, X. Zhao, and K. Huang. (2017b). "Locality-sensitive deconvolution networks with gated fusion for rgb-d indoor semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3029–3037.

Chiang, H.-Y., Y.-L. Lin, Y.-C. Liu, and W. H. Hsu. (2019). "A unified point-based framework for 3d segmentation". In: *2019 International Conference on 3D Vision (3DV)*. IEEE. 155–163.

Choy, C., J. Gwak, and S. Savarese. (2019). "4d spatio-temporal convnets: Minkowski convolutional neural networks". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3075–3084.

Chum, O. and A. Zisserman. (2007). "An exemplar model for learning object classes". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 1–8.

Comaniciu, M. d and P. M. Shift. (2002). "A Robust Approach toward Feature Space Analysis". *IEEE Trans. Patt. An. Mach. Intell.* 24.

Cordts, M., M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. (2016). "The cityscapes dataset for semantic urban scene understanding". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3213–3223.

Dai, A., A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner. (2017a). "Scannet: Richly-annotated 3d reconstructions of indoor scenes". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5828–5839.

Dai, A. and M. Nießner. (2018). "3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 452–468.

Dai, A., D. Ritchie, M. Bokeloh, S. Reed, J. Sturm, and M. Nießner. (2018). "Scancomplete: Large-scale scene completion and semantic segmentation for 3d scans". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4578–4587.

Dai, D., C. Sakaridis, S. Hecker, and L. Van Gool. (2020). "Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding". *International Journal of Computer Vision*. 128(5): 1182–1204.

Dai, D. and L. Van Gool. (2018). "Dark model adaptation: Semantic image segmentation from daytime to nighttime". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 3819–3824.

Dai, J., K. He, and J. Sun. (2016). "Instance-aware semantic segmentation via multi-task network cascades". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3150–3158.

Dai, J., H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. (2017b). "Deformable convolutional networks". In: *Proceedings of the IEEE international conference on computer vision*. 764–773.

Dai, Z., B. Cai, Y. Lin, and J. Chen. (2021). "Up-detr: Unsupervised pre-training for object detection with transformers". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1601–1610.

Dalal, N. and B. Triggs. (2005). "Histograms of oriented gradients for human detection". In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. Ieee. 886–893.

De Geus, D., P. Meletis, and G. Dubbelman. (2018). "Panoptic segmentation with a joint semantic and instance segmentation network". *arXiv preprint arXiv:1809.02110*.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. (2009). "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 248–255.

Deng, L., M. Yang, H. Li, T. Li, B. Hu, and C. Wang. (2019). "Restricted deformable convolution-based road scene semantic segmentation using surround view cameras". *IEEE Transactions on Intelligent Transportation Systems*. 21(10): 4350–4362.

Deng, L., M. Yang, Y. Qian, C. Wang, and B. Wang. (2017). "CNN based semantic segmentation for urban traffic scenes using fisheye camera". In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 231–236.

Ding, H., X. Jiang, B. Shuai, A. Q. Liu, and G. Wang. (2018). "Context contrasted feature and gated multi-scale aggregation for scene segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2393–2402.

Dong, B., F. Zeng, T. Wang, X. Zhang, and Y. Wei. (2021). "Solq: Segmenting objects by learning queries". *Advances in Neural Information Processing Systems*. 34: 21898–21909.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.* (2020). "An image is worth 16x16 words: Transformers for image recognition at scale". *arXiv preprint arXiv:2010.11929*.

Duan, K., S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian. (2019). "Centernet: Keypoint triplets for object detection". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6569–6578.

Dufour, R., C. Meurie, C. Strauss, and O. Lezoray. (2020). "Instance segmentation in fisheye images". In: *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE. 1–6.

Duke, B., A. Ahmed, C. Wolf, P. Aarabi, and G. W. Taylor. (2021). "Sstvos: Sparse spatiotemporal transformers for video object segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5912–5921.

Ehsani, K., R. Mottaghi, and A. Farhadi. (2018). "Segan: Segmenting and generating the invisible". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6144–6153.

Elsken, T., J. H. Metzen, and F. Hutter. (2019). "Neural architecture search: A survey". *The Journal of Machine Learning Research*. 20(1): 1997–2017.

Engelmann, F., T. Kontogianni, and B. Leibe. (2020). "Dilated point convolutions: On the receptive field size of point convolutions on 3d point clouds". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 9463–9469.

Erhan, D., C. Szegedy, A. Toshev, and D. Anguelov. (2014). "Scalable object detection using deep neural networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2147–2154.

Everingham, M., L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. (2010). "The pascal visual object classes (voc) challenge". *International journal of computer vision*. 88(2): 303–338.

Fan, S., Q. Dong, F. Zhu, Y. Lv, P. Ye, and F.-Y. Wang. (2021). "SCF-Net: Learning Spatial Contextual Features for Large-Scale Point Cloud Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14504–14513.

Fathi, A., Z. Wojna, V. Rathod, P. Wang, H. O. Song, S. Guadarrama, and K. P. Murphy. (2017). "Semantic instance segmentation via deep metric learning". *arXiv preprint arXiv:1703.10277*.

Ferrari, V., L. Fevrier, F. Jurie, and C. Schmid. (2007). "Groups of adjacent contour segments for object detection". *IEEE transactions on pattern analysis and machine intelligence*. 30(1): 36–51.

Fragkiadaki, K., P. Arbelaez, P. Felsen, and J. Malik. (2015). "Learning to segment moving objects in videos". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4083–4090.

Fu, H., Y. Xu, D. W. K. Wong, and J. Liu. (2016). "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields". In: *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*. IEEE. 698–701.

Fu, J., J. Liu, J. Jiang, Y. Li, Y. Bao, and H. Lu. (2020). "Scene segmentation with dual relation-aware attention network". *IEEE Transactions on Neural Networks and Learning Systems*. 32(6): 2547–2560.

Fu, J., J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu. (2019a). "Dual attention network for scene segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3146–3154.

Fu, J., J. Liu, Y. Wang, J. Zhou, C. Wang, and H. Lu. (2019b). "Stacked deconvolutional network for semantic segmentation". *IEEE Transactions on Image Processing*.

Gadde, R., V. Jampani, and P. V. Gehler. (2017). "Semantic video cnns through representation warping". In: *Proceedings of the IEEE International Conference on Computer Vision*. 4453–4462.

Gao, N., Y. Shan, X. Zhao, and K. Huang. (2021). "Learning Category- and Instance-Aware Pixel Embedding for Fast Panoptic Segmentation". *IEEE Transactions on Image Processing*. 30: 6013–6023. DOI: 10.1109/TIP.2021.3090522.

Gao, R. (2021). "Rethink Dilated Convolution for Real-time Semantic Segmentation". *arXiv preprint arXiv:2111.09957*.

Gasperini, S., M.-A. N. Mahani, A. Marcos-Ramiro, N. Navab, and F. Tombari. (2021). "Panoster: End-to-end panoptic segmentation of lidar point clouds". *IEEE Robotics and Automation Letters*. 6(2): 3216–3223.

Geiger, A., P. Lenz, and R. Urtasun. (2012). "Are we ready for autonomous driving? the kitti vision benchmark suite". In: *2012 IEEE conference on computer vision and pattern recognition*. IEEE. 3354–3361.

Gerke, M. (2014). "Use of the stair vision library within the ISPRS 2D semantic labeling benchmark (Vaihingen)".

Geus, D. de, P. Meletis, and G. Dubbelman. (2020). "Fast panoptic segmentation network". *IEEE Robotics and Automation Letters*. 5(2): 1742–1749.

Geus, D. de, P. Meletis, C. Lu, X. Wen, and G. Dubbelman. (2021). "Part-aware panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5485–5494.

Geyer, J., Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, *et al.* (2020). "A2d2: Audi autonomous driving dataset". *arXiv preprint arXiv:2004.06320*.

Girshick, R. (2015). "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

Girshick, R., J. Donahue, T. Darrell, and J. Malik. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

Golovin, D., B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley. (2017). "Google vizier: A service for black-box optimization". In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1487–1495.

Gong, K., X. Liang, D. Zhang, X. Shen, and L. Lin. (2017). "Look into person: Self-supervised structure-sensitive learning and a new benchmark for human parsing". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 932–940.

Gould, S., R. Fulton, and D. Koller. (2009). "Decomposing a scene into geometric and semantically consistent regions". In: *2009 IEEE 12th international conference on computer vision*. IEEE. 1–8.

Graber, C., G. Tsai, M. Firman, G. Brostow, and A. G. Schwing. (2021). "Panoptic segmentation forecasting". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12517–12526.

Graham, B., M. Engelcke, and L. Van Der Maaten. (2018). "3d semantic segmentation with submanifold sparse convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9224–9232.

Guo, R., D. Niu, L. Qu, and Z. Li. (2021). "SOTR: Segmenting Objects with Transformers". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7157–7166.

Gupta, A., P. Dollar, and R. Girshick. (2019). "LVIS: A dataset for large vocabulary instance segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5356–5364.

Gur, S., L. Wolf, L. Golgher, and P. Blinder. (2019). "Unsupervised microvascular image segmentation using an active contours mimicking neural network". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10722–10731.

Hackel, T., N. Savinov, L. Ladicky, J. Wegner, K. Schindler, and M. Pollefeys. (2017). "Semantic3D.net: A new Large-scale Point Cloud Classification Benchmark". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. IV-1/W1. DOI: 10.5194/isprs-annals-IV-1-W1-91-2017.

Haft-Javaherian, M., L. Fang, V. Muse, C. B. Schaffer, N. Nishimura, and M. R. Sabuncu. (2019). "Deep convolutional neural networks for segmenting 3D in vivo multiphoton images of vasculature in Alzheimer disease mouse models". *PloS one*. 14(3): e0213539.

Hahner, M., D. Dai, C. Sakaridis, J.-N. Zaech, and L. Van Gool. (2019). "Semantic understanding of foggy scenes with purely synthetic data". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE. 3675–3681.

Hao, S., Y. Zhou, and Y. Guo. (2020). "A brief survey on semantic segmentation with deep learning". *Neurocomputing*. 406: 302–321.

Hariharan, B., P. Arbeláez, R. Girshick, and J. Malik. (2014). "Simultaneous detection and segmentation". In: *European conference on computer vision*. Springer. 297–312.

Hatamizadeh, A., A. Hoogi, D. Sengupta, W. Lu, B. Wilcox, D. Rubin, and D. Terzopoulos. (2019a). "Deep active lesion segmentation". In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 98–105.

Hatamizadeh, A., D. Sengupta, and D. Terzopoulos. (2019b). "End-to-end deep convolutional active contours for image segmentation". *arXiv preprint arXiv:1909.13359*.

Hatamizadeh, A., D. Sengupta, and D. Terzopoulos. (2020). "End-to-end trainable deep active contour models for automated image segmentation: Delineating buildings in aerial imagery". In: *European Conference on Computer Vision*. Springer. 730–746.

Hayder, Z., X. He, and M. Salzmann. (2017). "Boundary-aware instance segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5696–5704.

He, J., Z. Deng, and Y. Qiao. (2019a). "Dynamic multi-scale filters for semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3562–3572.

He, J., Z. Deng, L. Zhou, Y. Wang, and Y. Qiao. (2019b). "Adaptive pyramid context network for semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7519–7528.

He, K., G. Gkioxari, P. Dollár, and R. Girshick. (2017). "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2961–2969.

He, K., X. Zhang, S. Ren, and J. Sun. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

Henry, P., M. Krainin, E. Herbst, X. Ren, and D. Fox. (2012). "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments". *The International Journal of Robotics Research*. 31(5): 647–663.

Hochreiter, S. and J. Schmidhuber. (1997). "Long short-term memory". *Neural computation*. 9(8): 1735–1780.

Hong, F., H. Zhou, X. Zhu, H. Li, and Z. Liu. (2021a). "Lidar-based panoptic segmentation via dynamic shifting network". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13090–13099.

Hong, W., Z. Wang, M. Yang, and J. Yuan. (2018). "Conditional generative adversarial network for structured domain adaptation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1335–1344.

Hong, Y., H. Pan, W. Sun, and Y. Jia. (2021b). "Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes". *arXiv preprint arXiv:2101.06085*.

Hoogi, A., A. Subramaniam, R. Veerapaneni, and D. L. Rubin. (2016). "Adaptive estimation of active contour parameters using convolutional neural networks and texture analysis". *IEEE transactions on medical imaging*. 36(3): 781–791.

Hosang, J., R. Benenson, P. Dollár, and B. Schiele. (2015). "What makes for effective detection proposals?" *IEEE transactions on pattern analysis and machine intelligence*. 38(4): 814–830.

Hou, R., J. Li, A. Bhargava, A. Raventos, V. Guizilini, C. Fang, J. Lynch, and A. Gaidon. (2020). "Real-Time Panoptic Segmentation From Dense Detections". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hu, J., L. Shen, and G. Sun. (2018a). "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7132–7141.

Hu, P., B. Shuai, J. Liu, and G. Wang. (2017). "Deep level sets for salient object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2300–2309.

Hu, Q., B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham. (2020). "Randla-net: Efficient semantic segmentation of large-scale point clouds". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11108–11117.

Hu, R., P. Dollár, K. He, T. Darrell, and R. Girshick. (2018b). "Learning to segment every thing". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4233–4241.

Hu, R., M. Rohrbach, and T. Darrell. (2016). "Segmentation from natural language expressions". In: *European Conference on Computer Vision*. Springer. 108–124.

Hu, X., K. Yang, L. Fei, and K. Wang. (2019). "Acnet: Attention based network to exploit complementary features for rgbd semantic segmentation". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 1440–1444.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. (2017). "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

Huang, H., Q. Huang, and P. Krahenbuhl. (2018a). "Domain transfer through deep activation matching". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 590–605.

Huang, J., D. Guan, A. Xiao, and S. Lu. (2021a). "Cross-View Regularization for Domain Adaptive Panoptic Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10133–10144.

Huang, J., D. Guan, A. Xiao, and S. Lu. (2021b). "Cross-view regularization for domain adaptive panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10133–10144.

Huang, J. and S. You. (2016). "Point cloud labeling using 3d convolutional neural network". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE. 2670–2675.

Huang, Q., W. Wang, and U. Neumann. (2018b). "Recurrent slice networks for 3d segmentation of point clouds". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2626–2635.

Huang, S., Z. Lu, R. Cheng, and C. He. (2021c). "FaPN: Feature-Aligned Pyramid Network for Dense Image Prediction". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 864–873.

Huang, X., P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang. (2019a). "The apolloscape open dataset for autonomous driving and its application". *IEEE transactions on pattern analysis and machine intelligence*. 42(10): 2702–2719.

Huang, P.-Y., W.-T. Hsu, C.-Y. Chiu, T.-F. Wu, and M. Sun. (2018c). "Efficient uncertainty estimation for semantic segmentation in videos". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 520–535.

Huang, Z., L. Huang, Y. Gong, C. Huang, and X. Wang. (2019b). "Mask scoring r-cnn". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6409–6418.

Huang, Z., X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu. (2019c). "Ccnet: Criss-cross attention for semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 603–612.

Hung, S.-W., S.-Y. Lo, and H.-M. Hang. (2019). "Incorporating luminance, depth and color information by a fusion-based network for semantic segmentation". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2374–2378.

Hung, W.-C., Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang. (2018). "Adversarial learning for semi-supervised semantic segmentation". *arXiv preprint arXiv:1802.07934*.

Hur, J. and S. Roth. (2016). "Joint optical flow and temporally consistent semantic segmentation". In: *European Conference on Computer Vision*. Springer. 163–177.

Ioffe, S. and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *International conference on machine learning*. PMLR. 448–456.

Jain, S. D., B. Xiong, and K. Grauman. (2017). "Fusionseg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos". In: *2017 IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE. 2117–2126.

Jaus, A., K. Yang, and R. Stiefelhagen. (2021). "Panoramic panoptic segmentation: Towards complete surrounding understanding via unsupervised contrastive learning". In: *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 1421–1427.

Jayasumana, S., K. Ranasinghe, M. Jayawardhana, S. Liyanaarachchi, and H. Ranasinghe. (2019). "Bipartite conditional random fields for panoptic segmentation". *arXiv preprint arXiv:1912.05307*.

Ji, X., J. F. Henriques, and A. Vedaldi. (2019). "Invariant information clustering for unsupervised image classification and segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9865–9874.

Jin, X., X. Li, H. Xiao, X. Shen, Z. Lin, J. Yang, Y. Chen, J. Dong, L. Liu, Z. Jie, *et al.* (2017). "Video scene parsing with predictive feature learning". In: *Proceedings of the IEEE International Conference on Computer Vision*. 5580–5588.

Kamnitsas, K., C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, *et al.* (2017). "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks". In: *International conference on information processing in medical imaging*. Springer. 597–609.

Kass, M., A. Witkin, and D. Terzopoulos. (1988). "Snakes: Active contour models". *International journal of computer vision*. 1(4): 321–331.

Kendall, A., V. Badrinarayanan, and R. Cipolla. (2017). "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding". DOI: 10.5244/C.31.57.

Khoreva, A., R. Benenson, J. Hosang, M. Hein, and B. Schiele. (2017). "Simple does it: Weakly supervised instance and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 876–885.

Khoreva, A., A. Rohrbach, and B. Schiele. (2019). "Video Object Segmentation with Language Referring Expressions": 123–141. DOI: 10.1007/978-3-030-20870-7_8.

Kim, D., S. Woo, J.-Y. Lee, and I. S. Kweon. (2020). "Video panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9859–9868.

Kirillov, A., R. Girshick, K. He, and P. Dollár. (2019a). "Panoptic feature pyramid networks". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6399–6408.

Kirillov, A., K. He, R. Girshick, C. Rother, and P. Dollár. (2019b). "Panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9404–9413.

Krähenbühl, P. and V. Koltun. (2011). "Efficient inference in fully connected crfs with gaussian edge potentials". *Advances in neural information processing systems*. 24: 109–117.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. (2012). "Imagenet classification with deep convolutional neural networks". *Advances in neural information processing systems*. 25: 1097–1105.

Kundu, A., V. Vineet, and V. Koltun. (2016). "Feature space optimization for semantic video segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3168–3175.

Landrieu, L. and M. Simonovsky. (2018). "Large-scale point cloud semantic segmentation with superpoint graphs". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4558–4567.

Lazarow, J., K. Lee, K. Shi, and Z. Tu. (2020). "Learning instance occlusion for panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10720–10729.

Le, T. H. N., K. G. Quach, K. Luu, C. N. Duong, and M. Savvides. (2018). "Reformulating level sets as deep recurrent neural network approach to semantic segmentation". *IEEE Transactions on Image Processing*. 27(5): 2393–2407.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner. (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*. 86(11): 2278–2324.

LeCun, Y., P. Haffner, L. Bottou, and Y. Bengio. (1999). "Object recognition with gradient-based learning". In: *Shape, contour and grouping in computer vision*. Springer. 319–345.

Lee, Y. and J. Park. (2020). "Centermask: Real-time anchor-free instance segmentation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 13906–13915.

Lengyel, A., S. Garg, M. Milford, and J. C. van Gemert. (2021). "Zero-Shot Day-Night Domain Adaptation with a Physics Prior". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4399–4409.

Li, H., P. Xiong, J. An, and L. Wang. (2018a). "Pyramid attention network for semantic segmentation". *arXiv preprint arXiv:1805.10180*.

Li, H., P. Xiong, H. Fan, and J. Sun. (2019a). "Dfanet: Deep feature aggregation for real-time semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9522–9531.

Li, H., G. Chen, G. Li, and Y. Yu. (2019b). "Motion guided attention for video salient object detection". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7274–7283.

Li, J., A. Zheng, X. Chen, and B. Zhou. (2017). "Primary video object segmentation via complementary CNNs and neighborhood reversible flow". In: *Proceedings of the IEEE international conference on computer vision*. 1417–1425.

Li, J., B. M. Chen, and G. H. Lee. (2018b). "So-net: Self-organizing network for point cloud analysis". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9397–9406.

Li, Q., A. Arnab, and P. H. Torr. (2018c). "Weakly- and Semi-Supervised Panoptic Segmentation". In: *Proceedings of the European Conference on Computer Vision (ECCV)*.

Li, Q., X. Qi, and P. H. Torr. (2020a). "Unifying training and inference for panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13320–13328.

Li, S., B. Seybold, A. Vorobyov, A. Fathi, Q. Huang, and C.-C. J. Kuo. (2018d). "Instance embedding transfer to unsupervised video object segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6526–6535.

Li, S., B. Seybold, A. Vorobyov, X. Lei, and C.-C. J. Kuo. (2018e). "Unsupervised video object segmentation with motion-based bilateral networks". In: *Proceedings of the European conference on computer vision (ECCV)*. 207–223.

Li, X., Z. Zhong, J. Wu, Y. Yang, Z. Lin, and H. Liu. (2019c). "Expectation-maximization attention networks for semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9167–9176.

Li, X., A. You, Z. Zhu, H. Zhao, M. Yang, K. Yang, S. Tan, and Y. Tong. (2020b). "Semantic flow for fast and accurate scene parsing". In: *European Conference on Computer Vision*. Springer. 775–793.

Li, X., L. Zhang, A. You, M. Yang, K. Yang, and Y. Tong. (2019d). "Global aggregation then local distribution in fully convolutional networks". *arXiv preprint arXiv:1909.07229*.

Li, X., H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng. (2018f). "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes". *IEEE transactions on medical imaging*. 37(12): 2663–2674.

Li, Y., X. Chen, Z. Zhu, L. Xie, G. Huang, D. Du, and X. Wang. (2019e). "Attention-guided unified network for panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7026–7035.

Li, Y., H. Zhao, X. Qi, L. Wang, Z. Li, J. Sun, and J. Jia. (2021). "Fully Convolutional Networks for Panoptic Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 214–223.

Li, Y., J. Moreau, and J. Ibanez-Guzman. (2022). "Unconventional Visual Sensors for Autonomous Vehicles". *arXiv preprint arXiv:2205.09383*.

Li, Y., J. Shi, and D. Lin. (2018g). "Low-latency video semantic segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5997–6005.

Li, Z., Y. Gan, X. Liang, Y. Yu, H. Cheng, and L. Lin. (2016). "Lstm-cf: Unifying context modeling and fusion with lstms for rgb-d scene labeling". In: *European conference on computer vision*. Springer. 541–557.

Liang, X., S. Liu, X. Shen, J. Yang, L. Liu, J. Dong, L. Lin, and S. Yan. (2015). "Deep human parsing with active template regression". *IEEE transactions on pattern analysis and machine intelligence*. 37(12): 2402–2414.

Liang, X., X. Shen, J. Feng, L. Lin, and S. Yan. (2016). "Semantic object parsing with graph lstm". In: *European Conference on Computer Vision*. Springer. 125–143.

Liao, S., Y. Gao, A. Oto, and D. Shen. (2013). "Representation learning: a unified deep learning framework for automatic prostate MR segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 254–261.

Lin, D., Y. Ji, D. Lischinski, D. Cohen-Or, and H. Huang. (2018). "Multi-scale context intertwining for semantic segmentation". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 603–619.

Lin, G., A. Milan, C. Shen, and I. Reid. (2017a). "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1925–1934.

Lin, G., C. Shen, A. Van Den Hengel, and I. Reid. (2016). "Efficient piecewise training of deep structured models for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3194–3203.

Lin, P., P. Sun, G. Cheng, S. Xie, X. Li, and J. Shi. (2020). "Graph-guided architecture search for real-time semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4203–4212.

Lin, T.-Y., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. (2017b). "Feature pyramid networks for object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2117–2125.

Lin, T.-Y., P. Goyal, R. Girshick, K. He, and P. Dollár. (2017c). "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2980–2988.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. (2014). "Microsoft coco: Common objects in context". In: *European conference on computer vision*. Springer. 740–755.

Litjens, G., R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, *et al.* (2014). "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge". *Medical image analysis*. 18(2): 359–373.

Liu, C., J. Yuen, and A. Torralba. (2011a). "Nonparametric scene parsing via label transfer". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 33(12): 2368–2382.

Liu, F., S. Li, L. Zhang, C. Zhou, R. Ye, Y. Wang, and J. Lu. (2017). "3DCNN-DQN-RNN: A deep reinforcement learning framework for semantic parsing of large-scale 3D point clouds". In: *Proceedings of the IEEE international conference on computer vision*. 5678–5687.

Liu, H., C. Peng, C. Yu, J. Wang, X. Liu, G. Yu, and W. Jiang. (2019a). "An end-to-end network for panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6172–6181.

Liu, H., J. Zhang, K. Yang, X. Hu, and R. Stiefelhagen. (2022). "CMX: Cross-Modal Fusion for RGB-X Semantic Segmentation with Transformers". *arXiv preprint arXiv:2203.04838*.

Liu, M.-Y., O. Tuzel, S. Ramalingam, and R. Chellappa. (2011b). "Entropy rate superpixel segmentation". In: *CVPR 2011*. IEEE. 2097–2104.

Liu, S., L. Qi, H. Qin, J. Shi, and J. Jia. (2018). "Path aggregation network for instance segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8759–8768.

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. (2016a). "Ssd: Single shot multibox detector". In: *European conference on computer vision*. Springer. 21–37.

Liu, W., A. Rabinovich, and A. Berg. (2016b). "ParseNet: Looking Wider to See Better". In:

Liu, X., Z. Han, Y.-S. Liu, and M. Zwicker. (2019b). "Point2sequence: Learning the shape representation of 3d point clouds with an attention-based sequence to sequence network". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 8778–8785.

Liu, Y., C. Shen, C. Yu, and J. Wang. (2020). "Efficient semantic video segmentation with per-frame inference". In: *European Conference on Computer Vision*. Springer. 352–368.

Liu, Y., B. Fan, S. Xiang, and C. Pan. (2019c). "Relation-shape convolutional neural network for point cloud analysis". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8895–8904.

Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. (2021). "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows". In: 9992–10002. DOI: 10.1109/ICCV48922.2021.00986.

Liu, Z., X. Li, P. Luo, C.-C. Loy, and X. Tang. (2015). "Semantic image segmentation via deep parsing network". In: *Proceedings of the IEEE international conference on computer vision*. 1377–1385.

Lo, S.-Y., H.-M. Hang, S.-W. Chan, and J.-J. Lin. (2019). "Efficient dense modules of asymmetric convolution for real-time semantic segmentation". In: *Proceedings of the ACM Multimedia Asia*. 1–6.

Long, J., E. Shelhamer, and T. Darrell. (2015). "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.

Lu, X., W. Wang, J. Shen, D. Crandall, and J. Luo. (2020). "Zero-shot video object segmentation with co-attention siamese networks". *IEEE transactions on pattern analysis and machine intelligence*.

Luc, P., C. Couprie, S. Chintala, and J. Verbeek. (2016). "Semantic segmentation using adversarial networks". *arXiv preprint arXiv:1611.08408*.

Luo, S., X.-C. Tai, L. Huo, Y. Wang, and R. Glowinski. (2019). "Convex shape prior for multi-object segmentation using a single level set function". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 613–621.

Luo, X., J. Zhang, K. Yang, A. Roitberg, K. Peng, and R. Stiefelhagen. (2022). "Towards Robust Semantic Segmentation of Accident Scenes via Multi-Source Mixed Sampling and Meta-Learning". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4429–4439.

Ma, C., J. Zhang, K. Yang, A. Roitberg, and R. Stiefelhagen. (2021). "Densepass: Dense panoramic semantic segmentation via unsupervised domain adaptation with attention-augmented context exchange". In: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE. 2766–2772.

Ma, L., J. Stückler, C. Kerl, and D. Cremers. (2017). "Multi-view deep learning for consistent semantic mapping with rgb-d cameras". In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 598–605.

Mahadevan, S., A. Athar, A. Ošep, S. Hennen, L. Leal-Taixé, and B. Leibe. (2020). "Making a case for 3D convolutions for object segmentation in Videos". *arXiv preprint arXiv:2008.11516*.

Mahasseni, B., S. Todorovic, and A. Fern. (2017). "Budget-aware deep semantic video segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1029–1038.

Malladi, R., J. A. Sethian, and B. C. Vemuri. (1995). "Shape modeling with front propagation: A level set approach". *IEEE transactions on pattern analysis and machine intelligence*. 17(2): 158–175.

Maninis, K.-K., S. Caelles, J. Pont-Tuset, and L. Van Gool. (2018). "Deep extreme cut: From extreme points to object segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 616–625.

Marcos, D., D. Tuia, B. Kellenberger, L. Zhang, M. Bai, R. Liao, and R. Urtasun. (2018). "Learning deep structured active contours end-to-end". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8877–8885.

Mason, J. C. and D. C. Handscomb. (2002). *Chebyshev polynomials*. CRC press.

Mehta, S., M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi. (2018). "Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation". In: *Proceedings of the european conference on computer vision (ECCV)*. 552–568.

Menze, B. H., A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.* (2014). "The multimodal brain tumor image segmentation benchmark (BRATS)". *IEEE transactions on medical imaging*. 34(10): 1993–2024.

Meyer, G. P., J. Charland, D. Hegde, A. Laddha, and C. Vallespi-Gonzalez. (2019). "Sensor fusion for joint 3d object detection and semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 0–0.

Milioto, A., J. Behley, C. McCool, and C. Stachniss. (2020). "LiDAR Panoptic Segmentation for Autonomous Driving". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 8505–8512. DOI: 10.1109/IROS45743.2020.9340837.

Milletari, F., N. Navab, and S.-A. Ahmadi. (2016). "V-net: Fully convolutional neural networks for volumetric medical image segmentation". In: *2016 fourth international conference on 3D vision (3DV)*. IEEE. 565–571.

Mo, K., S. Zhu, A. X. Chang, L. Yi, S. Tripathi, L. J. Guibas, and H. Su. (2019). "Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 909–918.

Mohajerani, S., R. Asad, K. Abhishek, N. Sharma, A. van Duynhoven, and P. Saeedi. (2019). "Cloudmaskgan: A content-aware unpaired image-to-image translation algorithm for remote sensing imagery". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 1965–1969.

Mohan, R. and A. Valada. (2021). "Efficientps: Efficient panoptic segmentation". *International Journal of Computer Vision*. 129(5): 1551–1579.

Moore, A. P., S. J. Prince, J. Warrell, U. Mohammed, and G. Jones. (2008). "Superpixel lattices". In: *2008 IEEE conference on computer vision and pattern recognition*. IEEE. 1–8.

Mottaghi, R., X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille. (2014). "The role of context for object detection and semantic segmentation in the wild". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 891–898.

Mozer, M. C. (1989). "A focused back-propagation algorithm for temporal pattern recognition". *Complex systems*. 3(4): 349–381.

Mumford, D. B. and J. Shah. (1989). "Optimal approximations by piecewise smooth functions and associated variational problems". *Communications on pure and applied mathematics*.

Nag, S., S. Adak, and S. Das. (2019). "What's there in the dark". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2996–3000.

Nekrasov, V., C. Shen, and I. Reid. (2018). "Light-weight refinenet for real-time semantic segmentation". *arXiv preprint arXiv:1810.03272*.

Neuhold, G., T. Ollmann, S. Rota Bulo, and P. Kontschieder. (2017). "The mapillary vistas dataset for semantic understanding of street scenes". In: *Proceedings of the IEEE international conference on computer vision*. 4990–4999.

Nguyen, A. and B. Le. (2013). "3D point cloud segmentation: A survey". In: *2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. 225–230. DOI: 10.1109/RAM.2013.6758588.

Nilsson, D. and C. Sminchisescu. (2018). "Semantic video segmentation by gated recurrent flow propagation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6819–6828.

Nirkin, Y., L. Wolf, and T. Hassner. (2021). "Hyperseg: Patch-wise hypernetwork for real-time semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4061–4070.

Noh, H., S. Hong, and B. Han. (2015). "Learning deconvolution network for semantic segmentation". In: *Proceedings of the IEEE international conference on computer vision*. 1520–1528.

O Pinheiro, P. O., R. Collobert, and P. Dollár. (2015). "Learning to segment object candidates". *Advances in neural information processing systems*. 28.

Orsic, M., I. Kreso, P. Bevandic, and S. Segvic. (2019). "In defense of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12607–12616.

Osher, S. and J. A. Sethian. (1988). "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations". *Journal of computational physics*. 79(1): 12–49.

Otsu, N. (1979). "A threshold selection method from gray-level histograms". *IEEE transactions on systems, man, and cybernetics*. 9(1): 62–66.

Ouahabi, A. and A. Taleb-Ahmed. (2021). "Deep learning for real-time semantic segmentation: Application in ultrasound imaging". *Pattern Recognition Letters*. 144: 27–34.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

Papandreou, G., L.-C. Chen, K. P. Murphy, and A. L. Yuille. (2015). "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation". In: *Proceedings of the IEEE international conference on computer vision*. 1742–1750.

Park, S.-J., K.-S. Hong, and S. Lee. (2017). "Rdfnet: Rgb-d multi-level residual feature fusion for indoor semantic segmentation". In: *Proceedings of the IEEE international conference on computer vision*. 4980–4989.

Paszke, A., A. Chaurasia, S. Kim, and E. Culurciello. (2016). "Enet: A deep neural network architecture for real-time semantic segmentation". *arXiv preprint arXiv:1606.02147*.

Peng, J., Y. Liu, S. Tang, Y. Hao, L. Chu, G. Chen, Z. Wu, Z. Chen, Z. Yu, Y. Du, *et al.* (2022a). "PP-LiteSeg: A Superior Real-Time Semantic Segmentation Model". *arXiv preprint arXiv:2204.02681*.

Peng, K., J. Fei, K. Yang, A. Roitberg, J. Zhang, F. Bieder, P. Heidenreich, C. Stiller, and R. Stiefelhagen. (2022b). "MASS: Multi-attentional semantic segmentation of LiDAR data for dense top-view understanding". *IEEE Transactions on Intelligent Transportation Systems*.

Peng, S., W. Jiang, H. Pi, X. Li, H. Bao, and X. Zhou. (2020). "Deep snake for real-time instance segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8533–8542.

Perazzi, F., J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung. (2016). "A benchmark dataset and evaluation methodology for video object segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 724–732.

Pham, Q.-H., B.-S. Hua, T. Nguyen, and S.-K. Yeung. (2019). "Real-time progressive 3D semantic segmentation for indoor scenes". In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 1089–1098.

Pinheiro, P. and R. Collobert. (2014). "Recurrent convolutional neural networks for scene labeling". In: *International conference on machine learning*. PMLR. 82–90.

Pinheiro, P. O., T.-Y. Lin, R. Collobert, and P. Dollár. (2016). "Learning to refine object segments". In: *European conference on computer vision*. Springer. 75–91.

Pont-Tuset, J., F. Perazzi, S. Caelles, P. Arbeláez, A. Sorkine-Hornung, and L. Van Gool. (2017). "The 2017 davis challenge on video object segmentation". *arXiv preprint arXiv:1704.00675*.

Porzi, L., S. R. Bulo, and P. Kontschieder. (2021). "Improving Panoptic Segmentation at All Scales". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7302–7311.

Qi, C. R., H. Su, K. Mo, and L. J. Guibas. (2017). "Pointnet: Deep learning on point sets for 3d classification and segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 652–660.

Qi, L., L. Jiang, S. Liu, X. Shen, and J. Jia. (2019). "Amodal instance segmentation with kins dataset". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3014–3023.

Qiao, S., Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen. (2021). "VIP-DeepLab: Learning Visual Perception With Depth-Aware Video Panoptic Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 3997–4008.

Qin, H., J. M. Zain, X. Ma, and T. Hai. (2010). "Scene segmentation based on seeded region growing for foreground detection". In: *2010 Sixth International Conference on Natural Computation*. Vol. 7. IEEE. 3619–3623.

Qiu, S., S. Anwar, and N. Barnes. (2021a). "Dense-resolution network for point cloud classification and segmentation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3813–3822.

Qiu, S., S. Anwar, and N. Barnes. (2021b). "Geometric back-projection network for point cloud classification". *IEEE Transactions on Multimedia*.

Qiu, S., S. Anwar, and N. Barnes. (2021c). "Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and Adaptive Fusion". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1757–1767.

Quan, T. M., D. G. Hildebrand, and W.-K. Jeong. (2016). "Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics".

Raj, A., D. Maturana, and S. Scherer. (2015). "Multi-scale convolutional architecture for semantic segmentation". *Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RITR-15-21*.

Ramli, M. F. and K. N. Tahar. (2020). "Homogeneous tree height derivation from tree crown delineation using Seeded Region Growing (SRG) segmentation". *Geo-spatial Information Science*. 23(3): 195–208.

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. (2016). "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

Ren, S., K. He, R. Girshick, and J. Sun. (2015). "Faster r-cnn: Towards real-time object detection with region proposal networks". *Advances in neural information processing systems*. 28: 91–99.

Ren, S., W. Liu, Y. Liu, H. Chen, G. Han, and S. He. (2021). "Reciprocal Transformations for Unsupervised Video Object Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15455–15464.

Riaz, H. U. M., N. Benbarka, and A. Zell. (2021). "FourierNet: Compact mask representation for instance segmentation using differentiable shape decoders". In: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 7833–7840.

Richter, S. R., Z. Hayder, and V. Koltun. (2017). "Playing for benchmarks". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2213–2222.

Richter, S. R., V. Vineet, S. Roth, and V. Koltun. (2016). "Playing for data: Ground truth from computer games". In: *European conference on computer vision*. Springer. 102–118.

Riegler, G., A. Osman Ulusoy, and A. Geiger. (2017). "Octnet: Learning deep 3d representations at high resolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3577–3586.

Romera, E., J. M. Alvarez, L. M. Bergasa, and R. Arroyo. (2017). "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation". *IEEE Transactions on Intelligent Transportation Systems*. 19(1): 263–272.

Romera, E., L. M. Bergasa, K. Yang, J. M. Alvarez, and R. Barea. (2019). "Bridging the day and night domain gap for semantic segmentation". In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 1312–1318.

Rong, W., Z. Li, W. Zhang, and L. Sun. (2014). "An improved CANNY edge detection algorithm". In: *2014 IEEE international conference on mechatronics and automation*. IEEE. 577–582.

Ronneberger, O., P. Fischer, and T. Brox. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 234–241.

Ros, G., S. Ramos, M. Granados, A. Bakhtiary, D. Vazquez, and A. M. Lopez. (2015). "Vision-based offline-online perception paradigm for autonomous driving". In: *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE. 231–238.

Ros, G., L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez. (2016). "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3234–3243.

Rother, C., V. Kolmogorov, and A. Blake. (2004). ""GrabCut" interactive foreground extraction using iterated graph cuts". *ACM transactions on graphics (TOG)*. 23(3): 309–314.

Rowley, H. A., S. Baluja, T. Kanade, *et al.* (1995). *Human face detection in visual scenes*. Citeseer.

Roy, A. and S. Todorovic. (2016). "A multi-scale cnn for affordance segmentation in rgb images". In: *European conference on computer vision*. Springer. 186–201.

Rupprecht, C., E. Huaroc, M. Baust, and N. Navab. (2016). "Deep active contours". *arXiv preprint arXiv:1607.05074*.

Sakaridis, C., D. Dai, and L. Gool. (2020). "Map-Guided Curriculum Domain Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PP: 1–1. DOI: 10.1109/TPAMI.2020.3045882.

Sakaridis, C., D. Dai, and L. V. Gool. (2019). "Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7374–7383.

Sakaridis, C., D. Dai, S. Hecker, and L. Van Gool. (2018a). "Model adaptation with synthetic and real data for semantic dense foggy scene understanding". In: *Proceedings of the european conference on computer vision (ECCV)*. 687–704.

Sakaridis, C., D. Dai, and L. Van Gool. (2018b). "Semantic foggy scene understanding with synthetic data". *International Journal of Computer Vision*. 126(9): 973–992.

Sankaranarayanan, S., Y. Balaji, A. Jain, S. N. Lim, and R. Chellappa. (2018). "Learning from synthetic data: Addressing domain shift for semantic segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3752–3761.

Saxena, A., S. H. Chung, A. Y. Ng, *et al.* (2005). "Learning depth from single monocular images". In: *NIPS*. Vol. 18. 1–8.

Shah, S., P. Ghosh, L. S. Davis, and T. Goldstein. (2018). "Stacked u-nets: a no-frills approach to natural image segmentation". *arXiv preprint arXiv:1804.10343*.

Shelhamer, E., K. Rakelly, J. Hoffman, and T. Darrell. (2016). "Clockwork convnets for video semantic segmentation". In: *European Conference on Computer Vision*. Springer. 852–868.

Shi, H., G. Lin, H. Wang, T.-Y. Hung, and Z. Wang. (2020). "Spsequencenet: Semantic segmentation network on 4d point clouds". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4574–4583.

Shi, J. and J. Malik. (2000). "Normalized cuts and image segmentation". *IEEE Transactions on pattern analysis and machine intelligence*. 22(8): 888–905.

Shin, H.-C., N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski. (2018). "Medical image synthesis for data augmentation and anonymization using generative adversarial networks". In: *International workshop on simulation and synthesis in medical imaging*. Springer. 1–11.

Shuai, B., Z. Zuo, B. Wang, and G. Wang. (2016). "Dag-recurrent neural networks for scene labeling". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3620–3629.

Silberman, N., D. Hoiem, P. Kohli, and R. Fergus. (2012). "Indoor segmentation and support inference from rgbd images". In: *European conference on computer vision*. Springer. 746–760.

Simonovsky, M. and N. Komodakis. (2017). "Dynamic edge-conditioned filters in convolutional neural networks on graphs". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3693–3702.

Simonyan, K. and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556*.

Song, H., W. Wang, S. Zhao, J. Shen, and K.-M. Lam. (2018). "Pyramid dilated deeper convlstm for video salient object detection". In: *Proceedings of the European conference on computer vision (ECCV)*. 715–731.

Song, S., S. P. Lichtenberg, and J. Xiao. (2015). "Sun rgb-d: A rgb-d scene understanding benchmark suite". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 567–576.

Song, S., F. Yu, A. Zeng, A. X. Chang, M. Savva, and T. Funkhouser. (2017). "Semantic scene completion from a single depth image". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1746–1754.

Souly, N., C. Spampinato, and M. Shah. (2017). "Semi supervised semantic segmentation using generative adversarial network". In: *Proceedings of the IEEE international conference on computer vision*. 5688–5696.

Staal, J., M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken. (2004). "Ridge-based vessel segmentation in color images of the retina". *IEEE transactions on medical imaging*. 23(4): 501–509.

Strudel, R., R. Garcia, I. Laptev, and C. Schmid. (2021). "Segmenter: Transformer for semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7262–7272.

Sun, L., J. Wang, K. Yang, K. Wu, X. Zhou, K. Wang, and J. Bai. (2021a). "Aerial-PASS: panoramic annular scene segmentation in drone videos". In: *2021 European Conference on Mobile Robots (ECMR)*. IEEE. 1–6.

Sun, L., K. Wang, K. Yang, and K. Xiang. (2019). "See clearer at night: towards robust nighttime semantic segmentation through day-night image conversion". In: *Artificial Intelligence and Machine Learning in Defense Applications*. Vol. 11169. International Society for Optics and Photonics. 111690A.

Sun, L., K. Yang, X. Hu, W. Hu, and K. Wang. (2020). "Real-time fusion network for RGB-D semantic segmentation incorporating unexpected obstacle detection for road-driving images". *IEEE Robotics and Automation Letters*. 5(4): 5558–5565.

Sun, P., R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, M. Tomizuka, L. Li, Z. Yuan, C. Wang, *et al.* (2021b). "Sparse r-cnn: End-to-end object detection with learnable proposals". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14454–14463.

Sun, Z., S. Cao, Y. Yang, and K. M. Kitani. (2021c). "Rethinking transformer-based set prediction for object detection". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3611–3620.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. (2015). "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

Takikawa, T., D. Acuna, V. Jampani, and S. Fidler. (2019). "Gated-scnn: Gated shape cnns for semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5229–5238.

Tao, A., K. Sapra, and B. Catanzaro. (2020). "Hierarchical multi-scale attention for semantic segmentation". *arXiv preprint arXiv:2005.10821*.

Teikari, P., M. Santos, C. Poon, and K. Hynynen. (2016). "Deep learning convolutional networks for multiphoton microscopy vasculature segmentation". *arXiv preprint arXiv:1606.02382*.

Thomas, H., C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas. (2019). "Kpconv: Flexible and deformable convolution for point clouds". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6411–6420.

Tian, Z., C. Shen, and H. Chen. (2020). "Conditional convolutions for instance segmentation". In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*. Springer. 282–298.

Tian, Z., C. Shen, H. Chen, and T. He. (2019). "Fcos: Fully convolutional one-stage object detection". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 9627–9636.

Tokmakov, P., K. Alahari, and C. Schmid. (2017). "Learning video object segmentation with visual memory". In: *Proceedings of the IEEE International Conference on Computer Vision*. 4481–4490.

Tran, D., L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. (2015). "Learning spatiotemporal features with 3d convolutional networks". In: *Proceedings of the IEEE international conference on computer vision*. 4489–4497.

Tran, D., L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. (2016). "Deep end2end voxel2voxel prediction". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 17–24.

Treml, M., J. Arjona-Medina, T. Unterthiner, R. Durgesh, F. Friedmann, P. Schuberth, A. Mayr, M. Heusel, M. Hofmarcher, M. Widrich, *et al.* (2016). "Speeding up semantic segmentation for autonomous driving".

Tsai, Y.-H., M.-H. Yang, and M. J. Black. (2016). "Video segmentation via object flow". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3899–3908.

Uijlings, J. R., K. E. Van De Sande, T. Gevers, and A. W. Smeulders. (2013). "Selective search for object recognition". *International journal of computer vision*. 104(2): 154–171.

Varma, G., A. Subramanian, A. Namboodiri, M. Chandraker, and C. Jawahar. (2019). "IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments". In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 1743–1751.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. (2017). "Attention is all you need". In: *Advances in neural information processing systems*. 5998–6008.

Vemulapalli, R., O. Tuzel, M.-Y. Liu, and R. Chellapa. (2016). "Gaussian conditional random field network for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3224–3233.

Visin, F., K. Kastner, K. Cho, M. Matteucci, A. Courville, and Y. Bengio. (2015). "A recurrent neural network based alternative to convolutional networks". *arXiv preprint arXiv:1505.00393*.

Visin, F., M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville. (2016). "Reseg: A recurrent neural network-based model for semantic segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 41–48.

Wang, H., W. Wang, and J. Liu. (2021a). "Temporal Memory Attention for Video Semantic Segmentation". In: 2254–2258. DOI: 10.1109/ICIP42928.2021.9506731.

Wang, H., R. Luo, M. Maire, and G. Shakhnarovich. (2020a). "Pixel Consensus Voting for Panoptic Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wang, H., Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen. (2021b). "MaX-DeepLab: End-to-End Panoptic Segmentation With Mask Transformers". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5463–5474.

Wang, H., Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen. (2020b). "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation". In: *European Conference on Computer Vision*. Springer. 108–126.

Wang, J., K. Yang, S. Gao, L. Sun, C. Zhu, K. Wang, and J. Bai. (2022). "High-performance panoramic annular lens design for real-time semantic segmentation on aerial imagery". *Optical Engineering*. 61(3): 035101.

Wang, J. and A. L. Yuille. (2015). "Semantic part segmentation using compositional model combining shape and appearance". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1788–1797.

Wang, P., P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell. (2018). "Understanding convolution for semantic segmentation". In: *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 1451–1460.

Wang, W., X. Lu, J. Shen, D. J. Crandall, and L. Shao. (2019a). "Zero-shot video object segmentation via attentive graph neural networks". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9236–9245.

Wang, W., T. Zhou, F. Porikli, D. Crandall, and L. Van Gool. (2021c). "A survey on deep learning technique for video segmentation". *arXiv preprint arXiv:2107.01153*.

Wang, X., T. Kong, C. Shen, Y. Jiang, and L. Li. (2020c). "Solo: Segmenting objects by locations". In: *European Conference on Computer Vision*. Springer. 649–665.

Wang, X., R. Zhang, T. Kong, L. Li, and C. Shen. (2020d). "Solov2: Dynamic and fast instance segmentation". *Advances in Neural information processing systems*. 33: 17721–17732.

Wang, Y., Q. Zhou, J. Liu, J. Xiong, G. Gao, X. Wu, and L. J. Latecki. (2019b). "Lednet: A lightweight encoder-decoder network for real-time semantic segmentation". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 1860–1864.

Wang, Y., Q. Zhou, J. Xiong, X. Wu, and X. Jin. (2019c). "Esnet: An efficient symmetric network for real-time semantic segmentation". In: *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer. 41–52.

Wang, Y., Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. (2019d). "Dynamic graph cnn for learning on point clouds". *Acm Transactions On Graphics (tog)*. 38(5): 1–12.

Wang, Z., D. Acuna, H. Ling, A. Kar, and S. Fidler. (2019e). "Object instance annotation with deep extreme level set evolution". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7500–7508.

Weber, M., J. Luiten, and B. Leibe. (2020). "Single-shot panoptic segmentation". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 8476–8483.

Wu, T., S. Tang, R. Zhang, J. Cao, and Y. Zhang. (2020a). "Cgnet: A light-weight context guided network for semantic segmentation". *IEEE Transactions on Image Processing*. 30: 1169–1179.

Wu, X., Z. Wu, H. Guo, L. Ju, and S. Wang. (2021). "Dannet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15769–15778.

Wu, Y., G. Zhang, Y. Gao, X. Deng, K. Gong, X. Liang, and L. Lin. (2020b). "Bidirectional graph reasoning network for panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9080–9089.

Wu, Y., G. Zhang, H. Xu, X. Liang, and L. Lin. (2020c). "Auto-Panoptic: Cooperative Multi-Component Architecture Search for Panoptic Segmentation". *arXiv preprint arXiv:2010.16119*.

Wu, Y., Y. Wu, G. Gkioxari, and Y. Tian. (2018). "Building generalizable agents with a realistic and rich 3d environment". *arXiv preprint arXiv:1801.02209*.

Wu, Y., A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. (2019). "Detectron2". URL: https://github.com/facebookresearch/detectron2.

Xia, F., P. Wang, X. Chen, and A. L. Yuille. (2017). "Joint multi-person pose estimation and semantic part segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6769–6778.

Xiang, K., K. Yang, and K. Wang. (2021). "Polarization-driven semantic segmentation via efficient attention-bridged fusion". *Optics Express*. 29(4): 4802–4820.

Xiang, Y. and D. Fox. (2017). "DA-RNN: Semantic Mapping with Data Associated Recurrent Neural Networks". DOI: 10.15607/RSS.2017.XIII.013.

Xie, E., P. Sun, X. Song, W. Wang, X. Liu, D. Liang, C. Shen, and P. Luo. (2020). "Polarmask: Single shot instance segmentation with polar representation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12193–12202.

Xie, E., W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. (2021). "SegFormer: Simple and efficient design for semantic segmentation with transformers". *Advances in Neural Information Processing Systems*. 34: 12077–12090.

Xie, S., R. Girshick, P. Dollár, Z. Tu, and K. He. (2017). "Aggregated residual transformations for deep neural networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1492–1500.

Xiong, Y., R. Liao, H. Zhao, R. Hu, M. Bai, E. Yumer, and R. Urtasun. (2019). "UPSNet: A Unified Panoptic Segmentation Network". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Xu, J., Z. Xiong, and S. P. Bhattacharyya. (2022). "PIDNet: A Real-time Semantic Segmentation Network Inspired from PID Controller". *arXiv preprint arXiv:2206.02066*.

Xu, M., Z. Zhang, H. Hu, J. Wang, L. Wang, F. Wei, X. Bai, and Z. Liu. (2021). "End-to-end semi-supervised object detection with soft teacher". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3060–3069.

Xu, N., B. Price, S. Cohen, and T. Huang. (2017). "Deep image matting". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2970–2979.

Xu, N., L. Yang, Y. Fan, D. Yue, Y. Liang, J. Yang, and T. Huang. (2018a). "Youtube-vos: A large-scale video object segmentation benchmark". *arXiv preprint arXiv:1809.03327*.

Xu, Y.-S., T.-J. Fu, H.-K. Yang, and C.-Y. Lee. (2018b). "Dynamic video segmentation network". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6556–6565.

Xu, W., H. Wang, F. Qi, and C. Lu. (2019). "Explicit shape encoding for real-time instance segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5168–5177.

Xue, Y., T. Xu, H. Zhang, L. R. Long, and X. Huang. (2018). "Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation". *Neuroinformatics*. 16(3): 383–392.

Yamaguchi, K., M. H. Kiapour, L. E. Ortiz, and T. L. Berg. (2012). "Parsing clothing in fashion photographs". In: *2012 IEEE Conference on Computer vision and pattern recognition*. IEEE. 3570–3577.

Yan, R., K. Yang, and K. Wang. (2021). "NLFNet: Non-Local Fusion Towards Generalized Multimodal Semantic Segmentation across RGB-Depth, Polarization, and Thermal Images". In: *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE. 1129–1135.

Yan, S., X.-C. Tai, J. Liu, and H.-Y. Huang. (2020a). "Convexity shape prior for level set-based image segmentation method". *IEEE Transactions on Image Processing*. 29: 7141–7152.

Yan, X., C. Zheng, Z. Li, S. Wang, and S. Cui. (2020b). "Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5589–5598.

Yang, C., X. Shi, D. Yao, and C. Li. (2017). "A level set method for convexity preserving segmentation of cardiac left ventricle". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2159–2163.

Yang, K., X. Hu, L. M. Bergasa, E. Romera, X. Huang, D. Sun, and K. Wang. (2019a). "Can we pass beyond the field of view? panoramic annular semantic segmentation for real-world surrounding perception". In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 446–453.

Yang, K., X. Hu, H. Chen, K. Xiang, K. Wang, and R. Stiefelhagen. (2020a). "Ds-pass: Detail-sensitive panoramic annular semantic segmentation through swaftnet for surrounding sensing". In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 457–464.

Yang, K., X. Hu, Y. Fang, K. Wang, and R. Stiefelhagen. (2020b). "Omnisupervised omnidirectional semantic segmentation". *IEEE Transactions on Intelligent Transportation Systems*.

Yang, K., X. Hu, and R. Stiefelhagen. (2021a). "Is context-aware cnn ready for the surroundings? panoramic semantic segmentation in the wild". *IEEE Transactions on Image Processing*. 30: 1866–1881.

Yang, K., J. Zhang, S. Reiß, X. Hu, and R. Stiefelhagen. (2021b). "Capturing omni-range context for omnidirectional segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1376–1386.

Yang, M., K. Yu, C. Zhang, Z. Li, and K. Yang. (2018). "Denseaspp for semantic segmentation in street scenes". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3684–3692.

Yang, Q., N. Li, Z. Zhao, X. Fan, E. Chang, and Y. Xu. (2020c). "MRI Cross-Modality Image-to-Image Translation". *Scientific Reports*. 10: 3753. DOI: 10.1038/s41598-020-60520-6.

Yang, S., Y. Fang, X. Wang, Y. Li, Y. Shan, B. Feng, and W. Liu. (2021c). "Tracking Instances as Queries". *arXiv preprint arXiv:2106.11963*.

Yang, T.-J., M. D. Collins, Y. Zhu, J.-J. Hwang, T. Liu, X. Zhang, V. Sze, G. Papandreou, and L.-C. Chen. (2019b). "Deeperlab: Single-shot image parser". *arXiv preprint arXiv:1902.05093*.

Yang, Y., H. Li, X. Li, Q. Zhao, J. Wu, and Z. Lin. (2020d). "Sognet: Scene overlap graph network for panoptic segmentation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 07. 12637–12644.

Ye, Y., K. Yang, K. Xiang, J. Wang, and K. Wang. (2020). "Universal semantic segmentation for fisheye urban driving images". In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 648–655.

Yi, L., V. G. Kim, D. Ceylan, I.-C. Shen, M. Yan, H. Su, C. Lu, Q. Huang, A. Sheffer, and L. Guibas. (2016). "A scalable active framework for region annotation in 3d shape collections". *ACM Transactions on Graphics (ToG)*. 35(6): 1–12.

Yin, M., Z. Yao, Y. Cao, X. Li, Z. Zhang, S. Lin, and H. Hu. (2020). "Disentangled non-local neural networks". In: *European Conference on Computer Vision*. Springer. 191–207.

Yu, B., L. Zhou, L. Wang, J. Fripp, and P. Bourgeat. (2018a). "3D cGAN based cross-modality MR image synthesis for brain tumor segmentation". In: 626–630. DOI: 10.1109/ISBI.2018.8363653.

Yu, C., J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang. (2018b). "Bisenet: Bilateral segmentation network for real-time semantic segmentation". In: *Proceedings of the European conference on computer vision (ECCV)*. 325–341.

Yu, C., J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang. (2018c). "Learning a discriminative feature network for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1857–1866.

Yu, F., H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell. (2020). "Bdd100k: A diverse driving dataset for heterogeneous multitask learning". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2636–2645.

Yu, F. and V. Koltun. (2015). "Multi-scale context aggregation by dilated convolutions". *arXiv preprint arXiv:1511.07122*.

Yuan, J., Z. Deng, S. Wang, and Z. Luo. (2020a). "Multi Receptive Field Network for Semantic Segmentation". In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 1883–1892.

Yuan, Y., X. Chen, X. Chen, and J. Wang. (2021a). "Segmentation transformer: Object-contextual representations for semantic segmentation". In: *European Conference on Computer Vision (ECCV)*. Vol. 1.

Yuan, Y., X. Chen, and J. Wang. (2020b). "Object-contextual representations for semantic segmentation". In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer. 173–190.

Yuan, Y., L. Huang, J. Guo, C. Zhang, X. Chen, and J. Wang. (2021b). "OCNet: Object Context for Semantic Segmentation". *International Journal of Computer Vision*. 129. DOI: 10.1007/s11263-021-01465-9.

Zagoruyko, S., A. Lerer, T.-Y. Lin, P. Pinheiro, S. Gross, S. Chintala, and P. Dollar. (2016). "A MultiPath Network for Object Detection". In: 15.1–15.12. DOI: 10.5244/C.30.15.

Zech, J. (2018). "What are radiological deep learning models actually learning". *Medium*. URL: https://medium.com/@jrzech/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98.

Zendel, O., K. Honauer, M. Murschitz, D. Steininger, and G. F. Dominguez. (2018). "Wilddash-creating hazard-aware benchmarks". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 402–416.

Zeng, A., K.-T. Yu, S. Song, D. Suo, E. Walker, A. Rodriguez, and J. Xiao. (2017). "Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge". In: *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE. 1386–1383.

Zhang, G., Y. Gao, H. Xu, H. Zhang, Z. Li, and X. Liang. (2020a). "Ada-Segment: Automated Multi-loss Adaptation for Panoptic Segmentation". *arXiv preprint arXiv:2012.03603*.

Zhang, G., J.-H. Xue, P. Xie, S. Yang, and G. Wang. (2021a). "Non-local aggregation for RGB-D semantic segmentation". *IEEE Signal Processing Letters*. 28: 658–662.

Zhang, H., K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal. (2018a). "Context encoding for semantic segmentation". In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 7151–7160.

Zhang, J., C. Ma, K. Yang, A. Roitberg, K. Peng, and R. Stiefelhagen. (2021b). "Transfer beyond the Field of View: Dense Panoramic Semantic Segmentation via Unsupervised Domain Adaptation". *IEEE Transactions on Intelligent Transportation Systems*.

Zhang, J., K. Yang, A. Constantinescu, K. Peng, K. Müller, and R. Stiefelhagen. (2021c). "Trans4Trans: Efficient Transformer for Transparent Object Segmentation To Help Visually Impaired People Navigate in the Real World". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*. 1760–1770.

Zhang, J., K. Yang, C. Ma, S. Reiß, K. Peng, and R. Stiefelhagen. (2022). "Bending reality: Distortion-aware transformers for adapting to panoramic semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16917–16927.

Zhang, J., K. Yang, and R. Stiefelhagen. (2021d). "Exploring Event-Driven Dynamic Context for Accident Scene Segmentation". *IEEE Transactions on Intelligent Transportation Systems*.

Zhang, L., J. Zhang, Z. Lin, R. Měch, H. Lu, and Y. He. (2020b). "Unsupervised video object segmentation with joint hotspot tracking". In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*. Springer. 490–506.

Zhang, R., S. A. Candra, K. Vetter, and A. Zakhor. (2015). "Sensor fusion for semantic segmentation of urban scenes". In: *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE. 1850–1857.

Zhang, R., Z. Tian, C. Shen, M. You, and Y. Yan. (2020c). "Mask encoding for single shot instance segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10226–10235.

Zhang, W., J. Pang, K. Chen, and C. C. Loy. (2021e). "K-Net: Towards Unified Image Segmentation". *arXiv preprint arXiv:2106.14855*.

Zhang, X., H. Xu, H. Mo, J. Tan, C. Yang, L. Wang, and W. Ren. (2021f). "Dcnas: Densely connected neural architecture search for semantic image segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13956–13967.

Zhang, Y., Z. Zhou, P. David, X. Yue, Z. Xi, B. Gong, and H. Foroosh. (2020d). "PolarNet: An Improved Grid Representation for Online LiDAR Point Clouds Semantic Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, Z., X. Zhang, C. Peng, X. Xue, and J. Sun. (2018b). "Exfuse: Enhancing feature fusion for semantic segmentation". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 269–284.

Zhang, Z., Z. Cui, C. Xu, Y. Yan, N. Sebe, and J. Yang. (2019). "Pattern-affinitive propagation across depth, surface normal and semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4106–4115.

Zhang, Z., L. Yang, and Y. Zheng. (2018c). "Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network". In: *Proceedings of the IEEE conference on computer vision and pattern Recognition*. 9242–9251.

Zhao, H., L. Jiang, J. Jia, P. H. Torr, and V. Koltun. (2021). "Point transformer". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 16259–16268.

Zhao, H., X. Qi, X. Shen, J. Shi, and J. Jia. (2018a). "Icnet for real-time semantic segmentation on high-resolution images". In: *Proceedings of the European conference on computer vision (ECCV)*. 405–420.

Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia. (2017). "Pyramid scene parsing network". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2881–2890.

Zhao, H., Y. Zhang, S. Liu, J. Shi, C. C. Loy, D. Lin, and J. Jia. (2018b). "Psanet: Point-wise spatial attention network for scene parsing". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 267–283.

Zhao, Y., T. Birdal, H. Deng, and F. Tombari. (2019). "3D point capsule networks". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1009–1018.

Zhen, M., S. Li, L. Zhou, J. Shang, H. Feng, T. Fang, and L. Quan. (2020). "Learning discriminative feature with crf for unsupervised video object segmentation". In: *European Conference on Computer Vision*. Springer. 445–462.

Zheng, M., P. Gao, X. Wang, H. Li, and H. Dong. (2020). "End-to-end object detection with adaptive clustering transformer". *arXiv preprint arXiv:2011.09315*.

Zheng, S., S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr. (2015). "Conditional random fields as recurrent neural networks". In: *Proceedings of the IEEE international conference on computer vision*. 1529–1537.

Zheng, S., J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. Torr, and L. Zhang. (2021). "Rethinking Semantic Segmentation From a Sequence-to-Sequence Perspective With Transformers". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 6881–6890.

Zhong, Z., Z. Q. Lin, R. Bidart, X. Hu, I. B. Daya, Z. Li, W.-S. Zheng, J. Li, and A. Wong. (2020). "Squeeze-and-attention networks for semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13065–13074.

Zhou, B., H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba. (2017). "Scene parsing through ade20k dataset". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 633–641.

Zhou, T., S. Wang, Y. Zhou, Y. Yao, J. Li, and L. Shao. (2020). "Motion-attentive transition for zero-shot video object segmentation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 07. 13066–13073.

Zhou, X., J. Zhuo, and P. Krahenbuhl. (2019). "Bottom-up object detection by grouping extreme and center points". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 850–859.

Zhou, Z., Y. Zhang, and H. Foroosh. (2021). "Panoptic-PolarNet: Proposal-Free LiDAR Point Cloud Panoptic Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13194–13203.

Zhu, X., H. Zhou, T. Wang, F. Hong, Y. Ma, W. Li, H. Li, and D. Lin. (2021). "Cylindrical and asymmetrical 3d convolution networks for lidar segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9939–9948.

Zhu, X., W. Su, L. Lu, B. Li, X. Wang, and J. Dai. (2020). "Deformable detr: Deformable transformers for end-to-end object detection". *arXiv preprint arXiv:2010.04159*.

Zhu, X., Y. Xiong, J. Dai, L. Yuan, and Y. Wei. (2017a). "Deep feature flow for video recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2349–2358.

Zhu, Y., Y. Tian, D. Metaxas, and P. Dollár. (2017b). "Semantic amodal segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1464–1472.

Zhuang, J., J. Yang, L. Gu, and N. Dvornek. (2019). "Shelfnet for fast semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 0–0.

Zitnick, C. L. and P. Dollár. (2014). "Edge boxes: Locating object proposals from edges". In: *European conference on computer vision*. Springer. 391–405.

Zoph, B., G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. Le. (2020). "Rethinking pre-training and self-training". *Advances in neural information processing systems*. 33: 3833–3845.

Zoph, B. and Q. V. Le. (2016). "Neural architecture search with reinforcement learning". *arXiv preprint arXiv:1611.01578*.