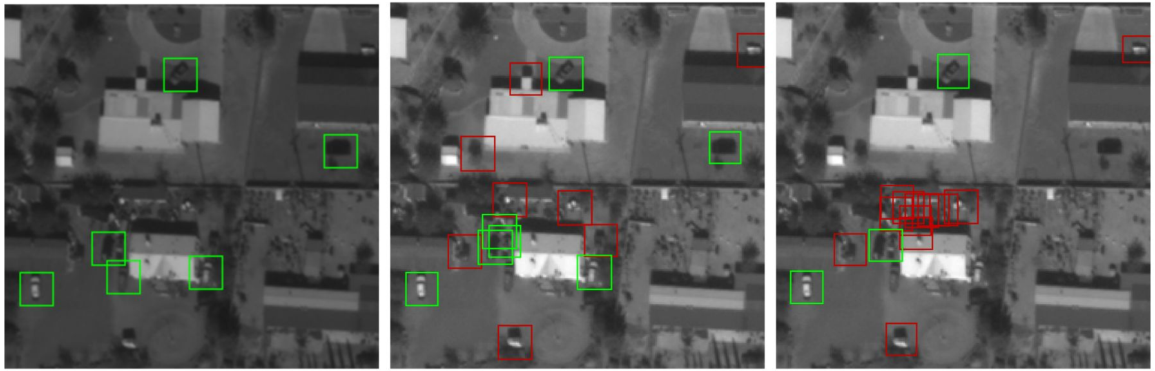
CiPSI and CiPSINet

A New Dataset and Network for Finding Cars in Panchromatic Satellite Images

Benjamin Parfitt - Mapware - BenjaminLParfitt@gmail.com

Tej Gagadam - Mapware - tejaishwarya.gagadam@mapware.ai



(a) Ground Truth (b) Faster R-CNN CiPSINet (c) Faster R-CNN ResNet-50

ICMLA 2021

Mapware

Overview

1. Dataset
2. Architecture
3. Classification
4. Detection
5. Future Work

Motivation

1. There are gaps in existing datasets for the classification and detection of cars in black and white satellite images.
2. Existing methods often use higher resolution images that are downsampled captured from drones or planes.

Problems with the drone/plane-based approach:

- No atmospheric distortions
- No wide field of view
- No cloud cover

CiPSI Dataset

The images are from the Worldview 3 and Worldview 4 satellites, and were provided by the European Space Agency. Locations include landlocked and coastal areas, and have a GSD of ~35cm.

Detection:

- Each scene was annotated by two people to ensure that all cars were found.

Classification:

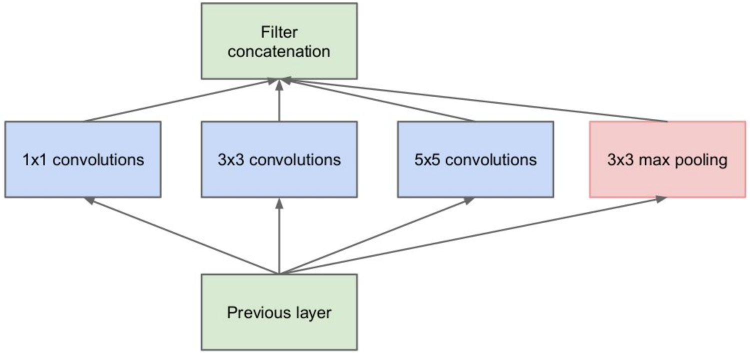
- Each annotated car was clipped from the scene, including surrounding context (Mundhenk et al, 2016).
- Non-car regions were clipped from these scenes as well.

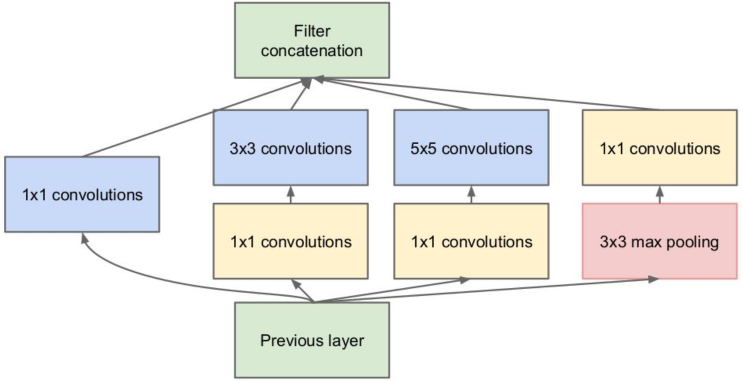
T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, “A large contextual dataset for classification, detection and counting of cars with deep learning,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9907 LNCS, pp. 785–800, 2016.

Existing Deep Learning Architectures

Inception

- Inception Blocks





C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-December, pp. 2818–2826, 2016.

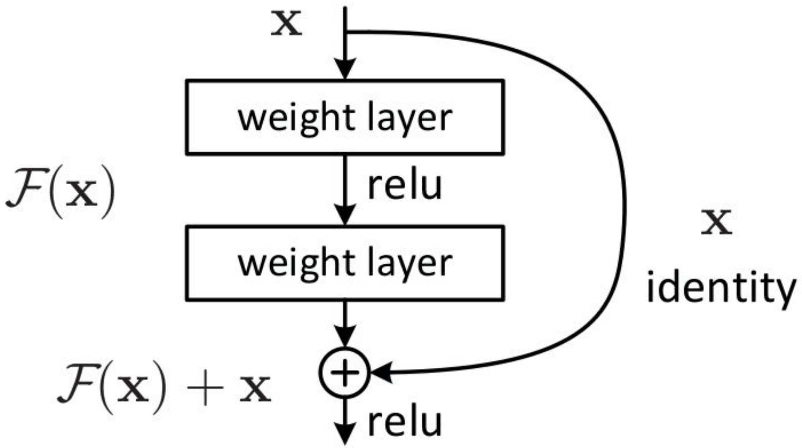
Existing Deep Learning Architectures

Inception

- Inception Blocks

ResNet and Wide ResNet

- Residuals
- Width vs Depth



S. Zagoruyko and N. Komodakis, “Wide residual networks,” British Machine Vision Conference 2016, BMVC 2016, vol. 2016-September, pp. 87.1–87.12, 2016.

Existing Deep Learning Architectures

Inception

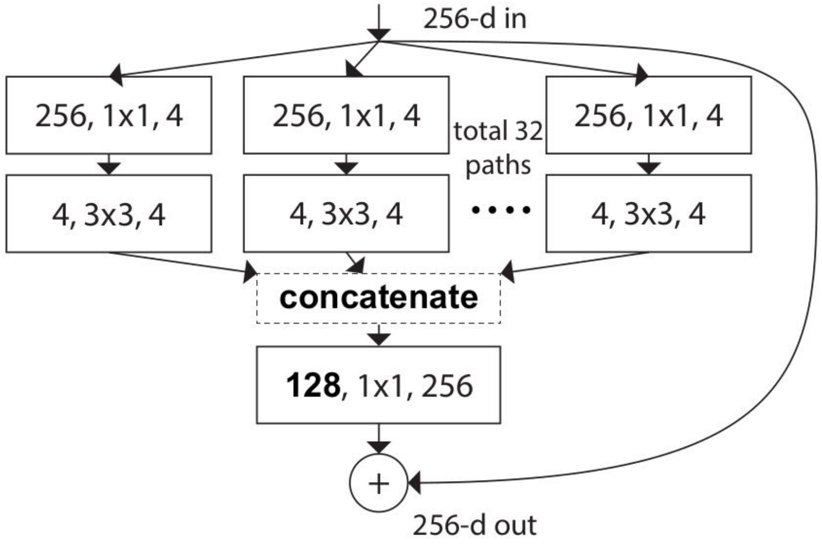
- Inception Blocks

ResNet and Wide ResNet

- Residuals
- Width vs Depth

ResNext

- Cardinality



S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 5987–5995, 2017.

The CiPSINet Architecture

The main techniques:

- Increase cardinality
- Limit width to keep costs low
- Limit filter size to keep costs low
- Limit depth to keep costs low - only 14 layers

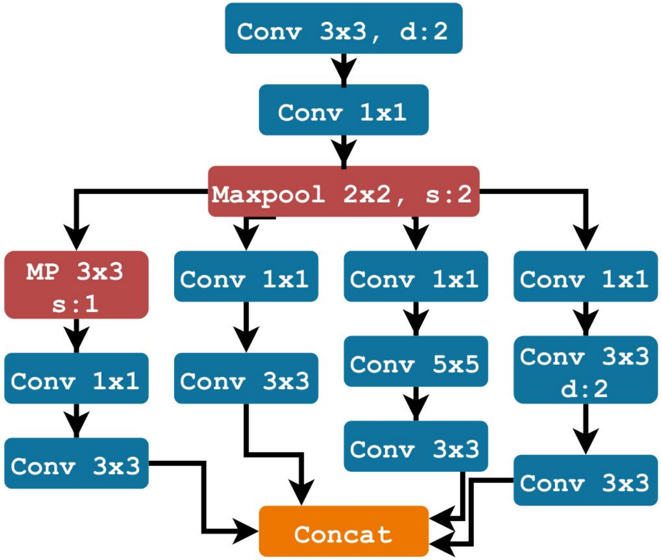
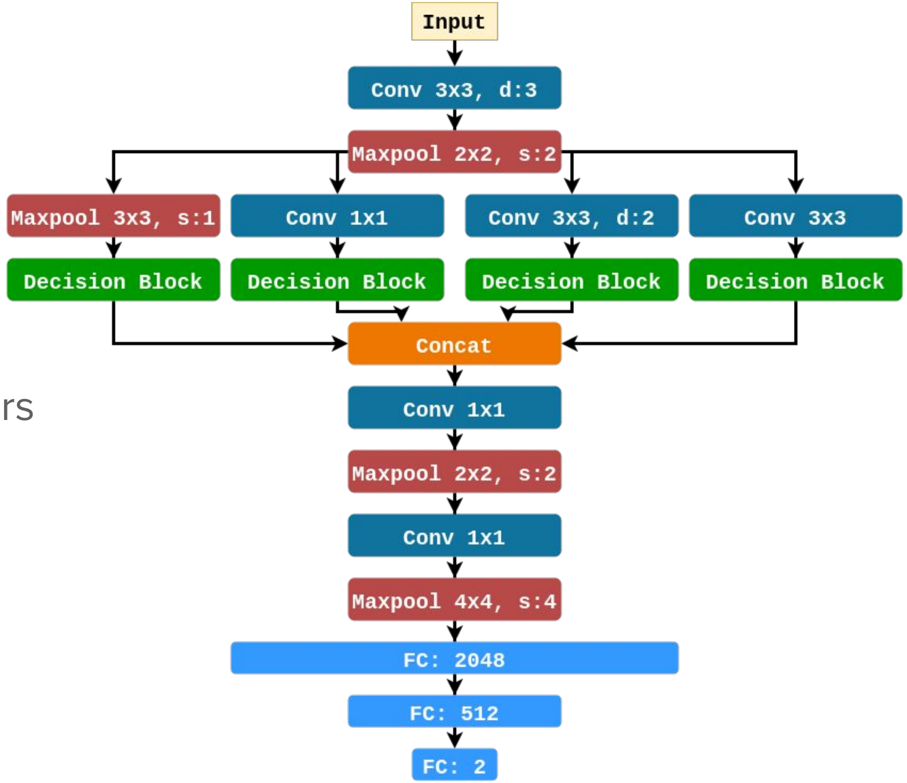


Fig. 1: The DecisionBlock architecture. s indicates stride, d indicates dilation. Both are 1 if not indicated. MP is MaxPool, Conv is convolutional, and Concat is Concatenate.

The CiPSINet Architecture

The main techniques:

- Increase cardinality
- Limit width to keep costs low
- Limit filter size to keep costs low
- Limit depth to keep costs low - only 14 layers



Classification Results

Results on test data after each epoch

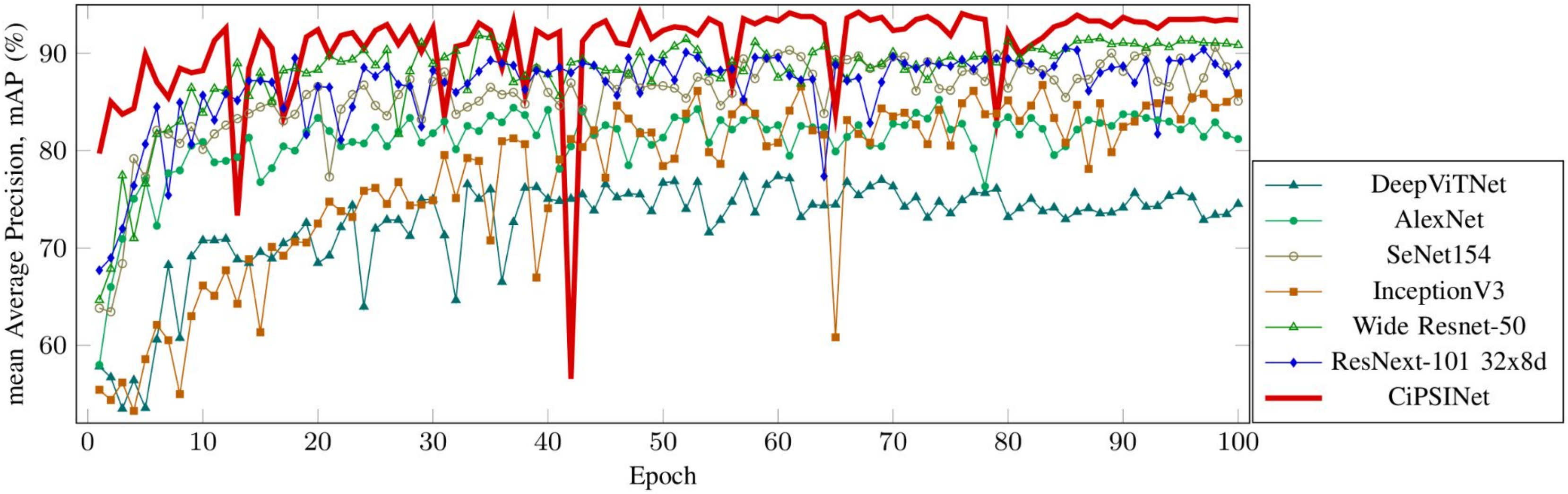
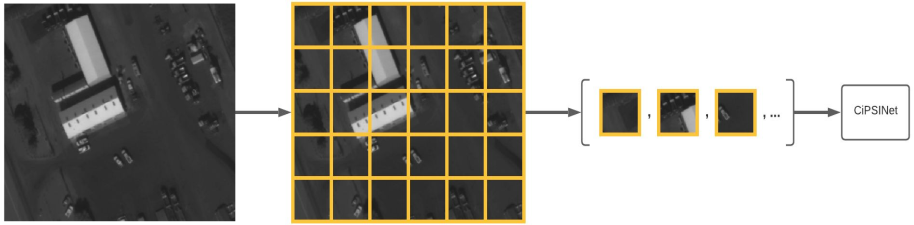


Fig. 2: The mAP of each network on the test data at the end of every epoch during training.

Detection Methods

Detection methods chosen for comparison:

1. **Sliding Window**
   - Fixed-length window moves across the image
   - CiPSINet classifies each cell as car or no car



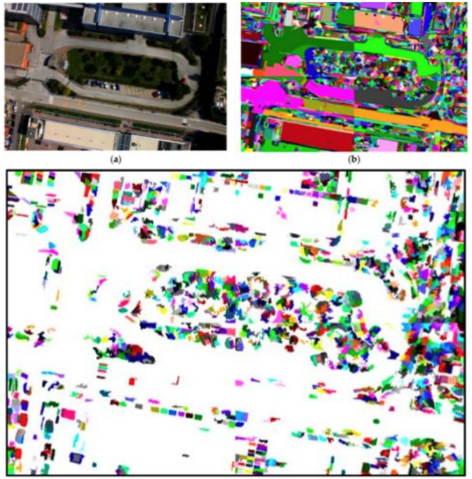
2. Unsupervised-Segmentation-Generated Region Proposals
3. Faster R-CNN

T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, “A large contextual dataset for classification, detection and counting of cars with deep learning,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9907 LNCS, pp. 785–800, 2016.
S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” june 2015. [Online].

Detection Methods

Detection methods chosen for comparison:

1. Sliding Window
2. **Unsupervised-Segmentation-Generated Region Proposals**
   - Mean-shift segmentation
   - Classifying this using CiPSINet
3. Faster R-CNN



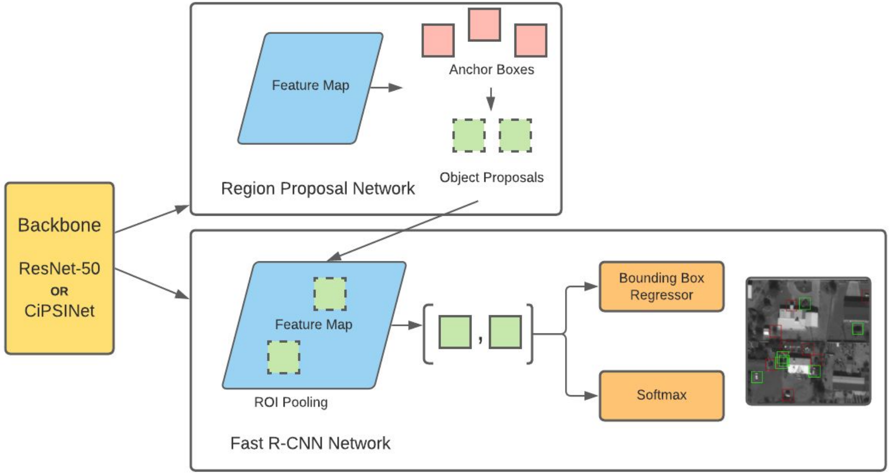
Regions obtained from mean-shift segmentation filtering

N. Ammour et al., “Deep learning approach for car detection in UAV imagery,” Remote Sensing, vol. 9, no. 4, 2017.

Detection Methods

Faster R-CNN with varying backbones:

- With Resnet-50 pre-trained on COCO
- With CiPSINet pre-trained on car classification



Faster R-CNN with ResNet-50 or CiPSINet backbone

S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," june 2015. [Online]. Available: http://arxiv.org/abs/1506.01497
S. M. Azimi, R. Bahmanyar, C. Henry, and F. Kurz, "EAGLE: Large-scale vehicle detection dataset in real-world scenarios using aerial imagery," pp. 6920–6927, 2021

Detection Results

Test Results				
Method	mAP @ 0.5	mAP @ [0.5, 0.95]	mAR @ [0.5, 0.95]	Time per Scene (ms)
F. R-CNN with CiPSINet	**49.3**	**26.1**	**45.6**	98.36
F. R-CNN with ResNet-50	36.2	21.4	42.5	**34.49**
Segmentation	3	1.2	34	2156.26
Sliding Window	< 0.1	< 0.1	1	12181.38

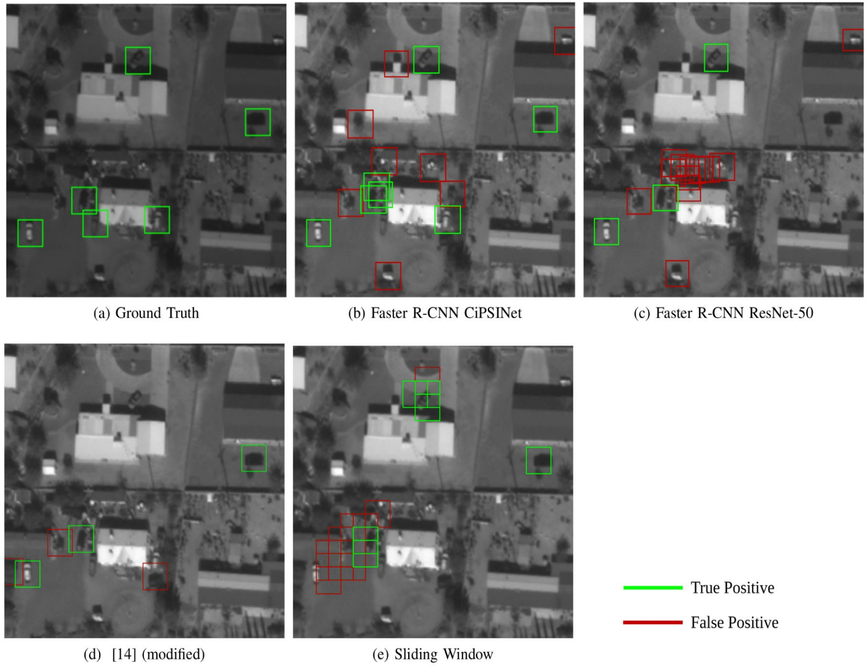


Fig. 3: Results of detection on a test scene. The ground truth is indicated in (a) by green boxes. In all other scenes the green boxes represent correctly identified cars, and the red boxes represent incorrectly identified cars.

Detection Results

- False positives identified by Faster R-CNN with CiPSINet and ResNet-50 as the backbone (Ref: images (a) and (b))

- Faster R-CNN with CiPSINet backbone identifies partially hidden cars as compared to Faster R-CNN with ResNet-50 (Ref: image (c) and (d))

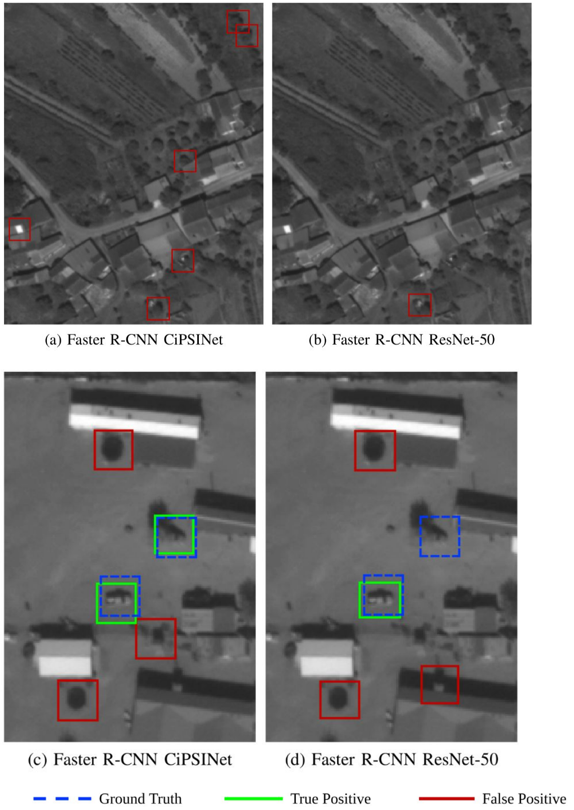


Fig. 4: Results of detection on two test scenes by Faster R-CNN with either CiPSINet or ResNet-50 as the backbone network.

Future Work

1. Dataset
   a. Continue to expand with further challenging examples (e.g. partially occluded cars)
2. Architecture
   a. Evaluate the performance on other benchmark datasets
   b. Ablation studies
   c. Continue to reduce the parameter count
      i. Reducing fully connected layer size
      ii. Using nx1 -> 1xn convolutional filters (n x n > n + n for n > 1)

Questions

Thank you for your time!

Existing Deep Learning Architectures

Inception

- Inception Blocks

ResNet and Wide ResNet

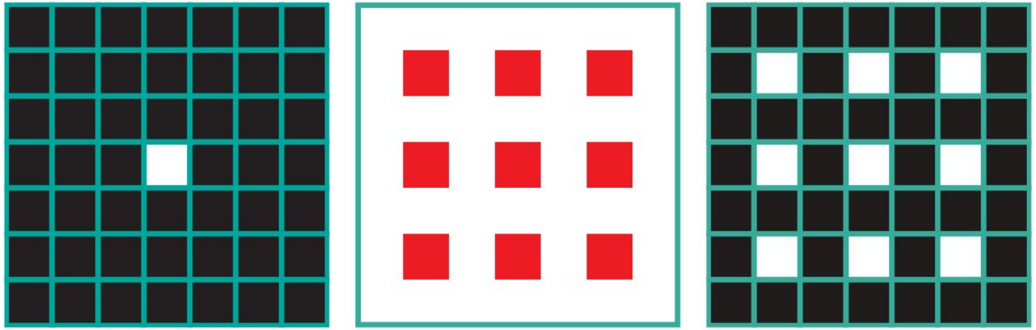
- Residuals
- Width vs Depth

ResNext

- Cardinality

Dilated Residual Networks

- Dilation



V. Koltun, F. Yu, and T. Funkhouser, “Dilated residual networks,” Cvpr, 2017. [Online]. Available: https://arxiv.org/pdf/1705.09914.pdf