Google   JOHNS HOPKINS   Stanford University
                    UNIVERSITY

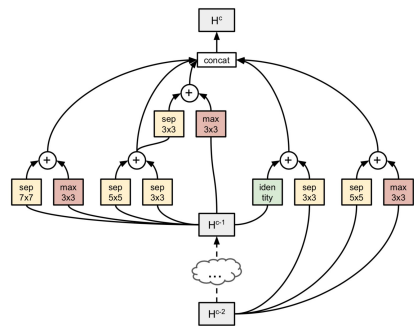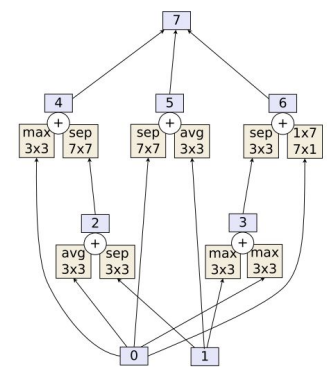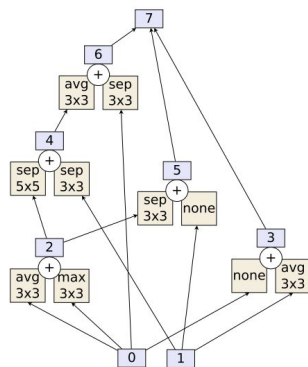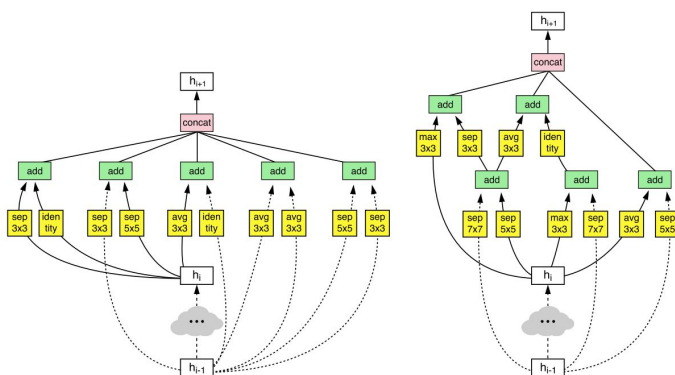# Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation

Chenxi Liu, Liang-Chieh Chen, Florian Schroff, Hartwig Adam, Wei Hua, Alan Yuille, Li Fei-Fei

06/18/2019 @CVPR

Neural Architecture Search for Image Classification
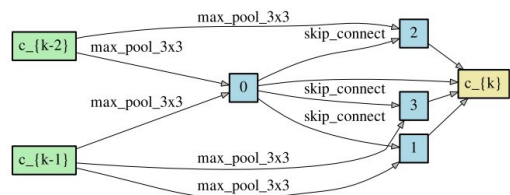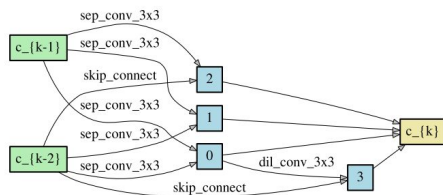
Zoph, Barret, et al. "Learning transferable architectures for scalable image recognition." In CVPR. 2018.
Liu, Chenxi, et al. "Progressive neural architecture search." In ECCV. 2018.
Real, Esteban, et al. "Regularized evolution for image classifier architecture search." In AAAI. 2019.
Liu, Hanxiao, Karen Simonyan, and Yiming Yang. "Darts: Differentiable architecture search." In ICLR. 2019.

Neural Architecture Search for Dense Image Prediction

- Image classification is a good starting point for NAS, but should not be the end point.

- Our paper is one of the first efforts to extend NAS to dense image prediction (semantic segmentation to be exact).

Challenge 1: Network Level Search Space
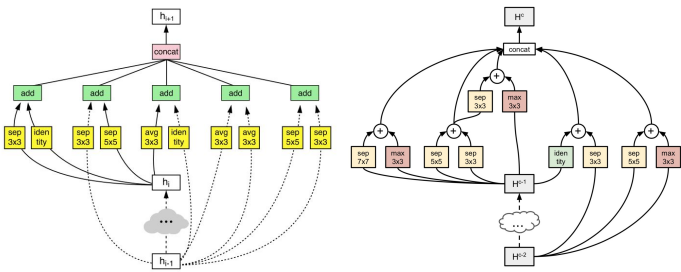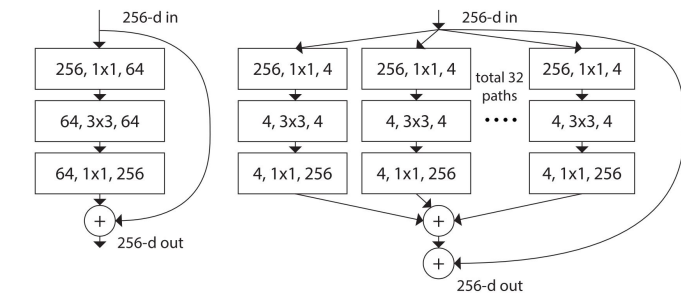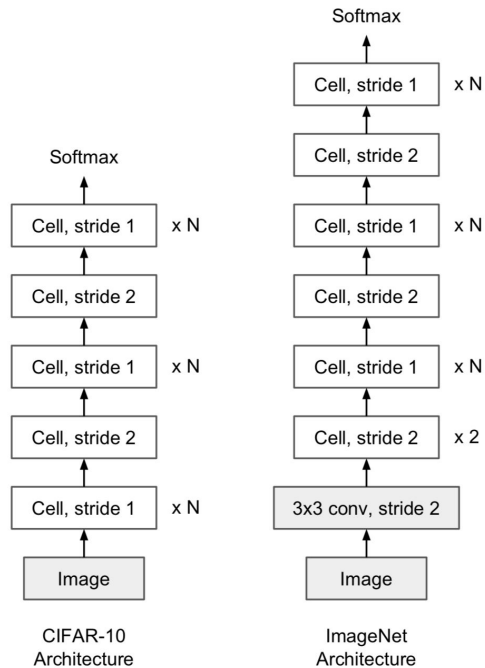
Inner Cell Level

Outer Network Level

Challenge 1: Network Level Search Space
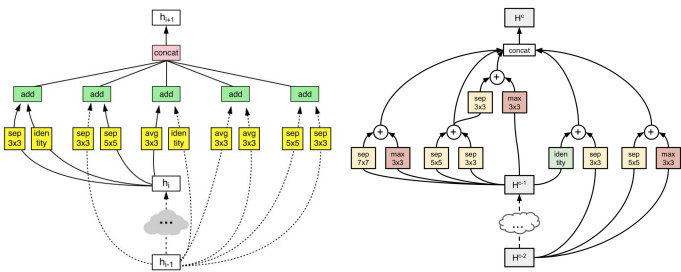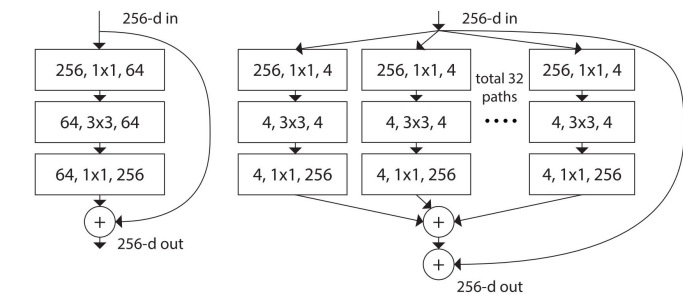
Inner Cell Level
(automatically search)
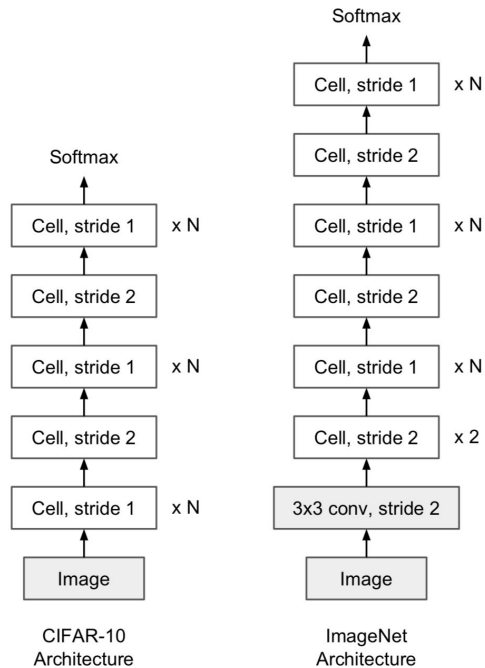
Outer Network Level
(hand design)

Challenge 2: Need for High Resolution & Efficient NAS

Challenge 2: Need for High Resolution & Efficient NAS

32x32

→

airplane

Challenge 2: Need for High Resolution & Efficient NAS

32x32

airplane

> 321x321
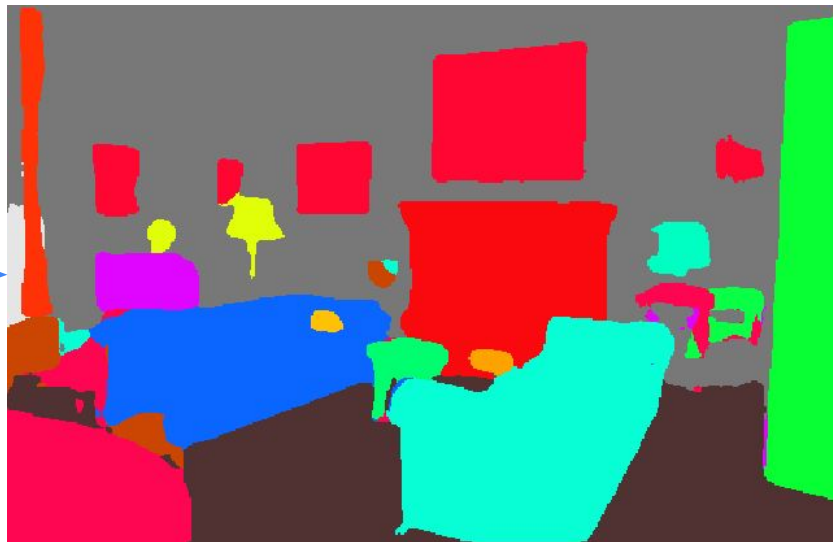
Idea of Differentiable NAS
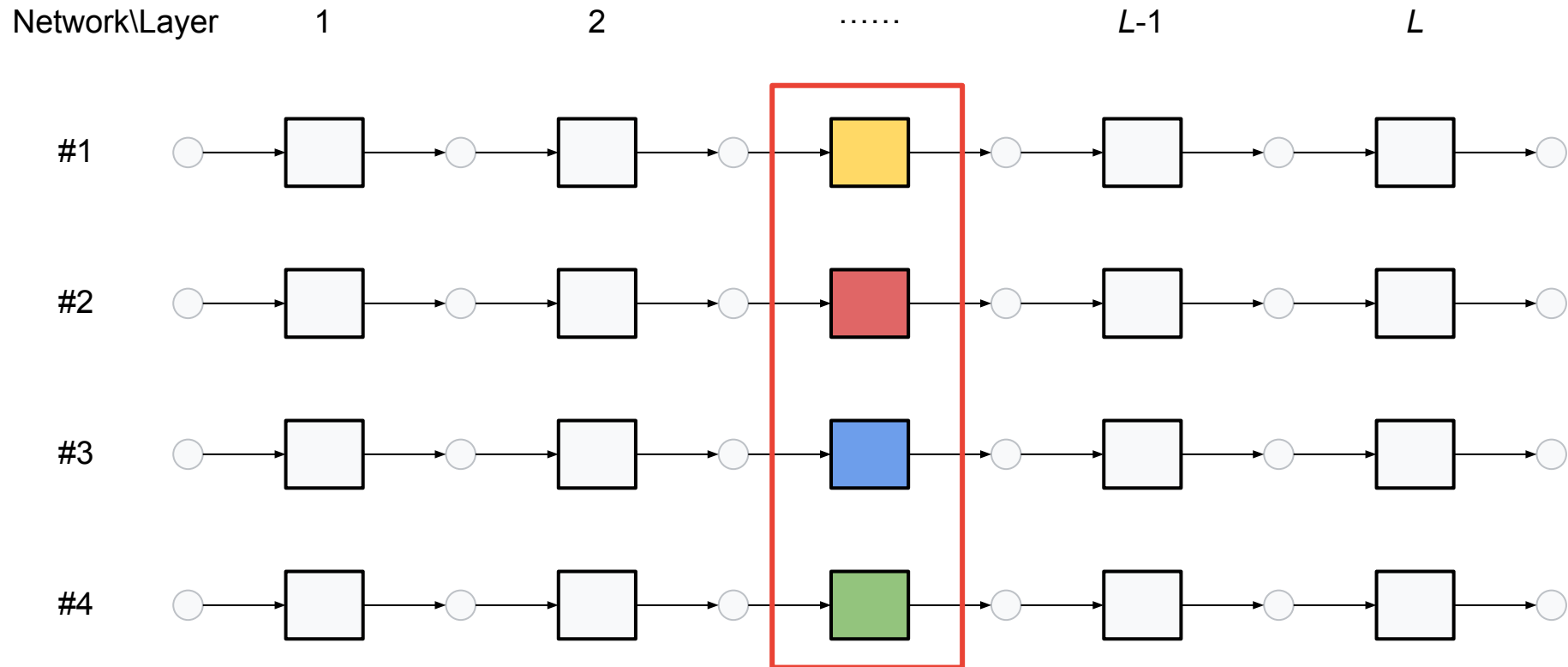
Idea of Differentiable NAS
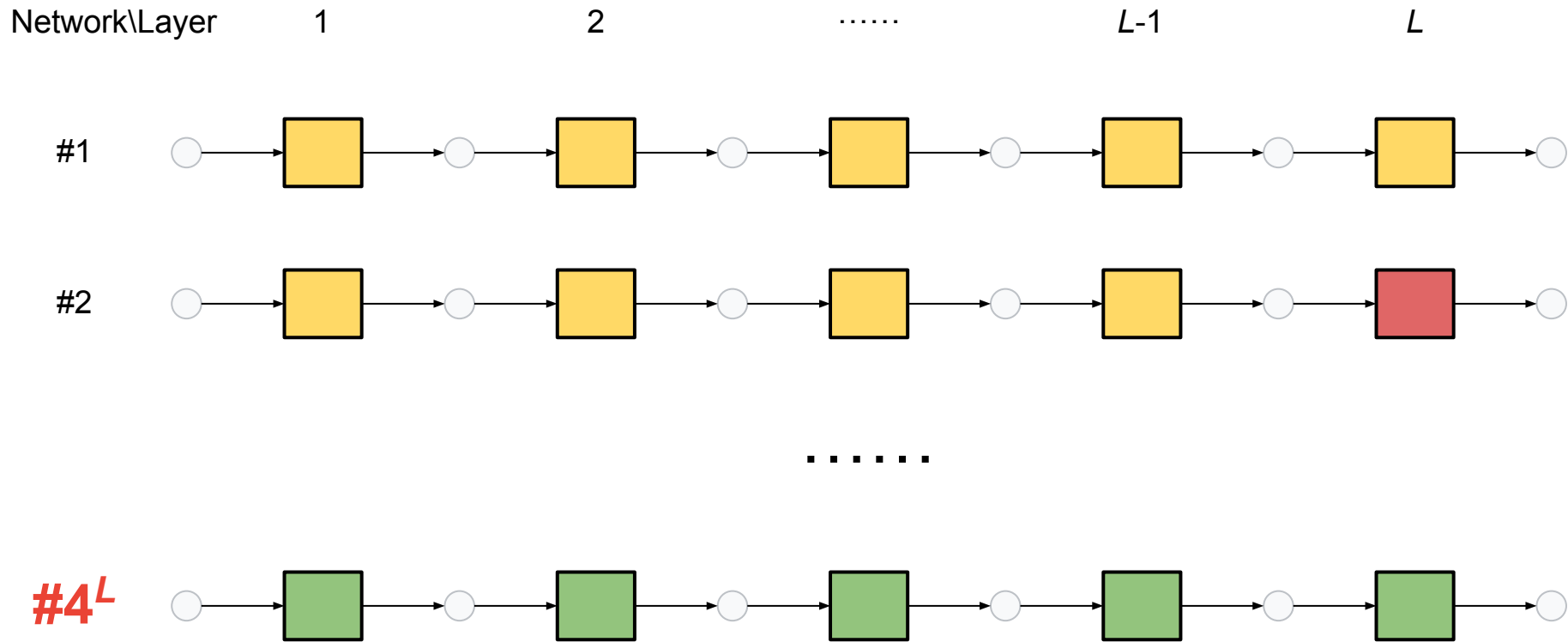
Idea of Differentiable NAS

Network\Layer   1   2   ......   *L*-1   *L*

#1

Idea of Differentiable NAS

Network\Layer   1   2   …… L-1 L

Liu, Hanxiao, Karen Simonyan, and Yiming Yang. "Darts: Differentiable architecture search." In ICLR. 2019.

Idea of Differentiable NAS

Network\Layer   1   2   …… L-1 L

α_3 is the largest among the four

#1

Liu, Hanxiao, Karen Simonyan, and Yiming Yang. "Darts: Differentiable architecture search." In ICLR. 2019.
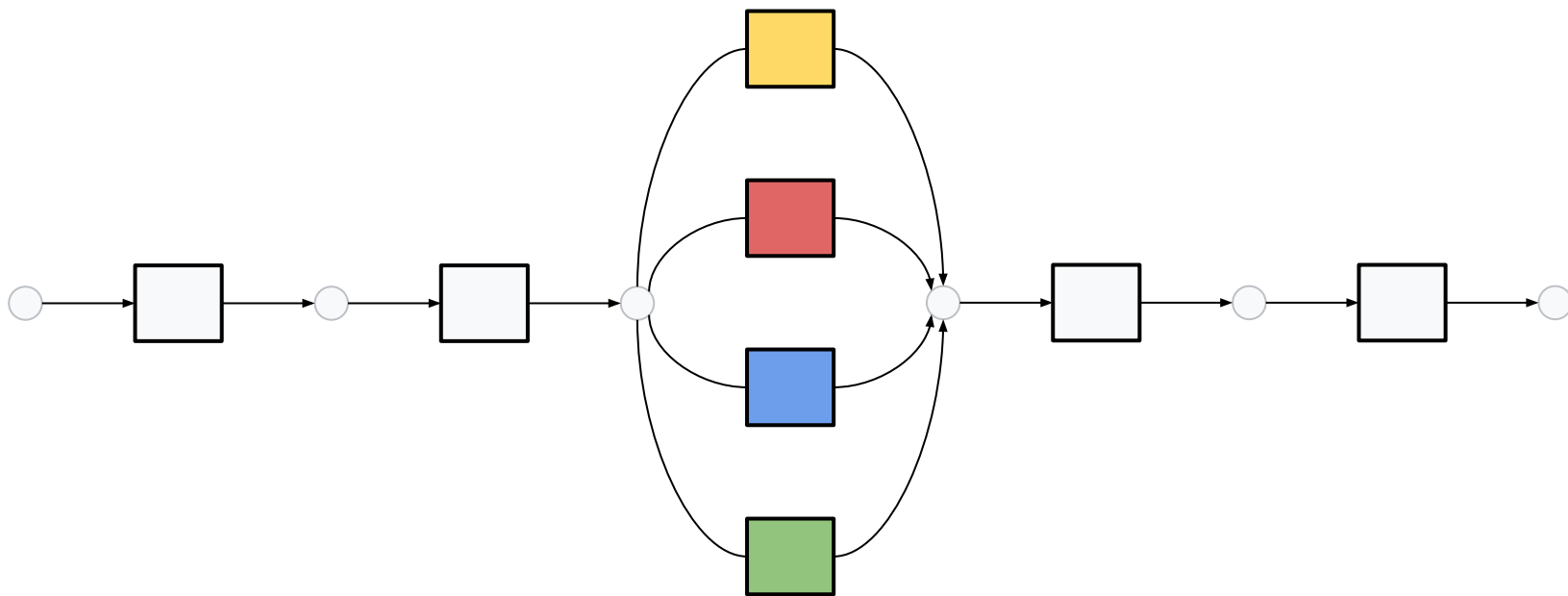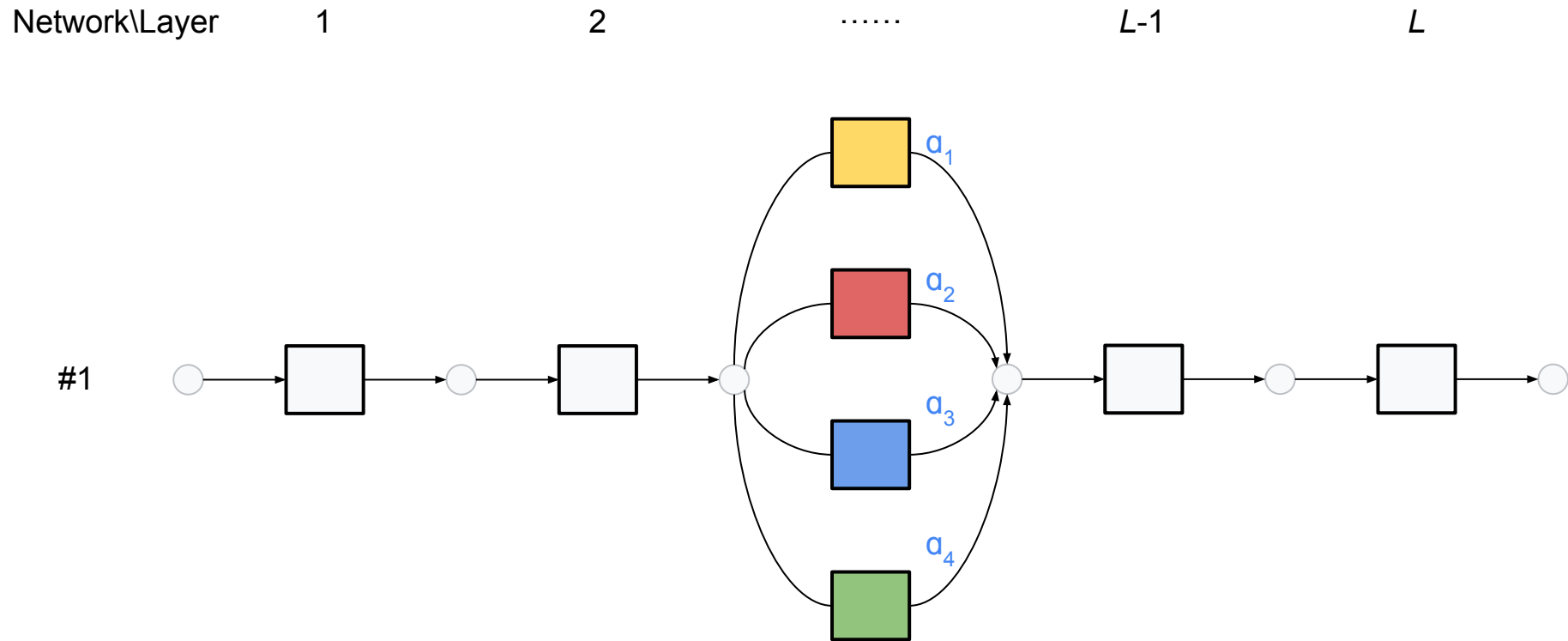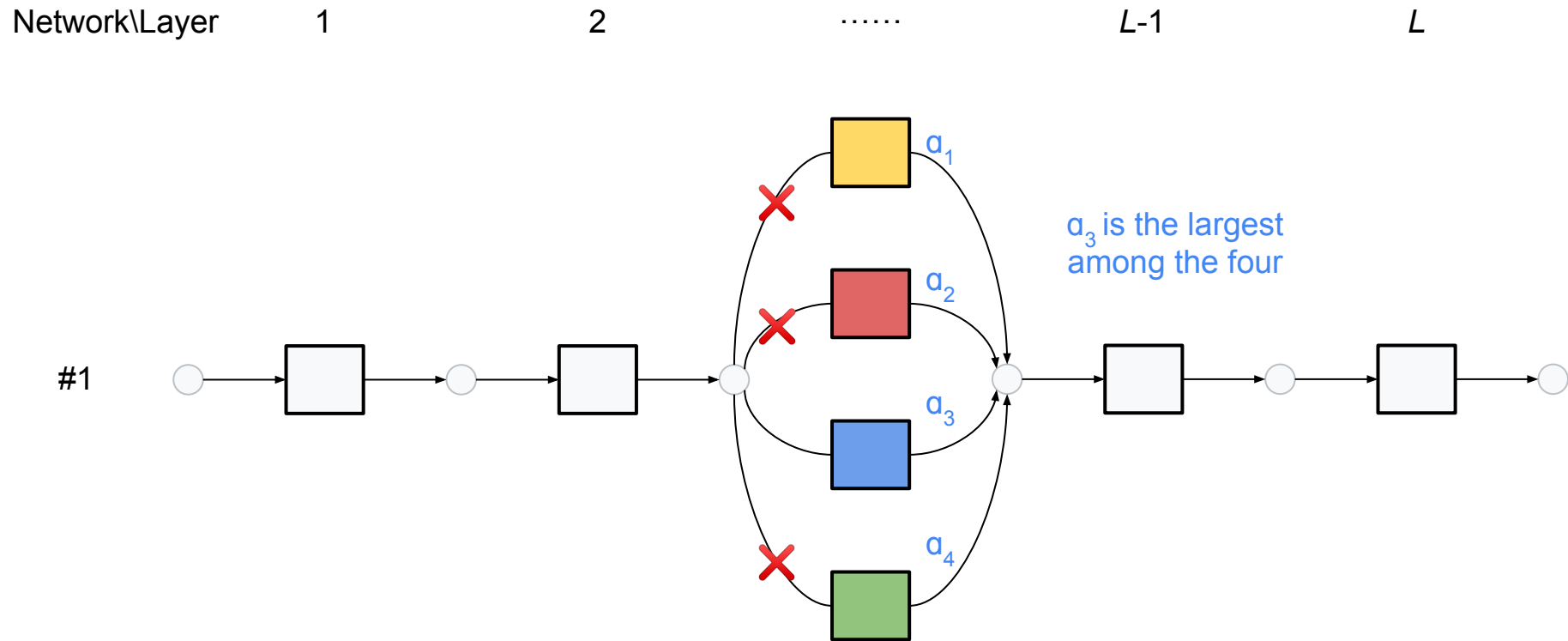
Idea of Differentiable NAS

Network\Layer   1   2   …… L-1 L

#1

Network Level Search Space

Downsample\Layer	1	2	3	4	5	……	L-1	L
1								
2								
4								
8								
16								
……								

Network Level Search Space

Downsample\Layer	1	2	3	4	5	......	L-1	L
1								
2								
4								
8								
16								

......

Network Level Search Space

Downsample\Layer	1	2	3	4	5	……	$L-1$	L
1								
2								
4								
8								
16								
……								

Network Level Search Space

Downsample\Layer	1	2	3	4	5		......		*L-1*	*L*
1										
2										
4										
8										
16										
......										

Network Level Search Space

Downsample\Layer	1	2	3	4	5	……	L-1	L
1								
2								
4								
8								
16								
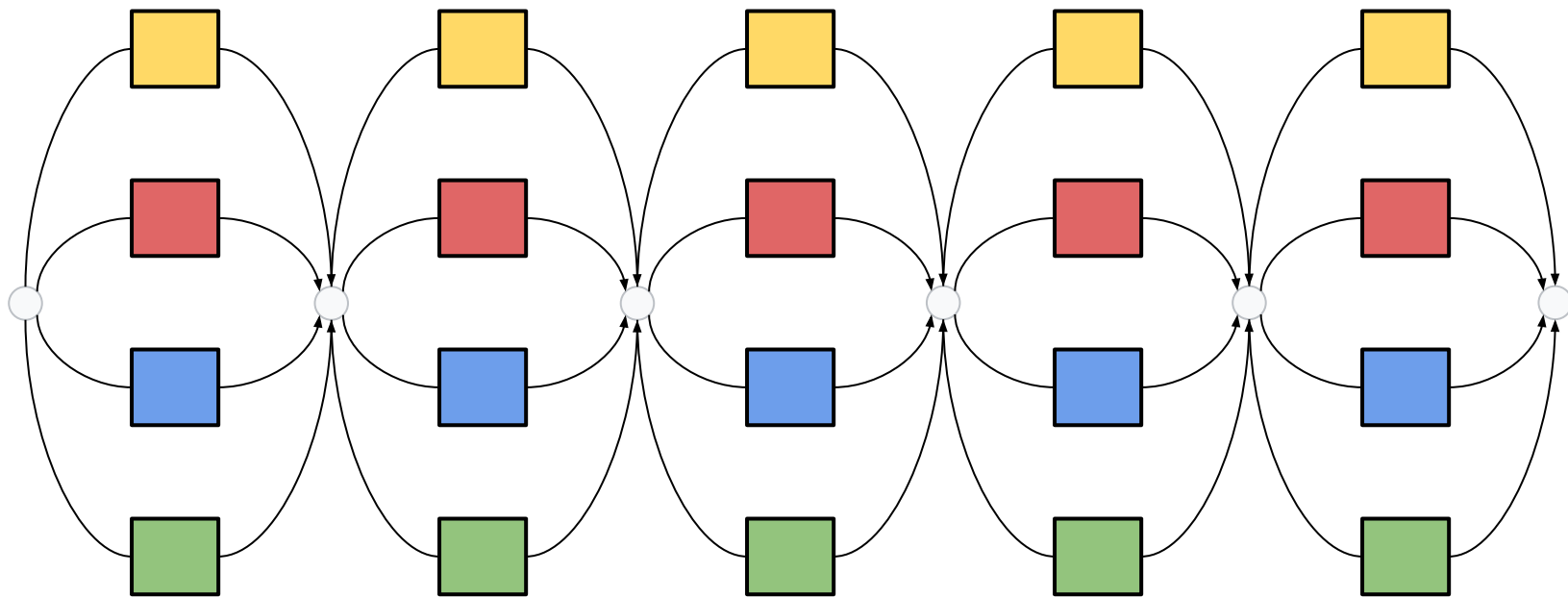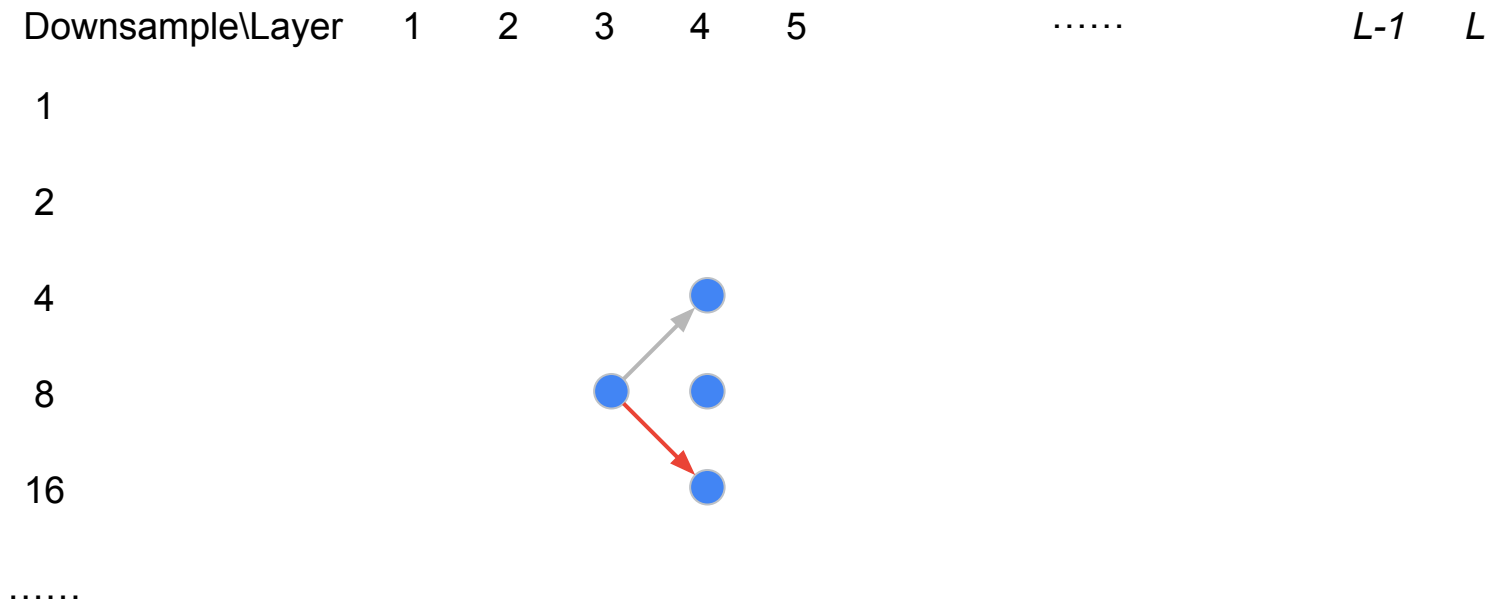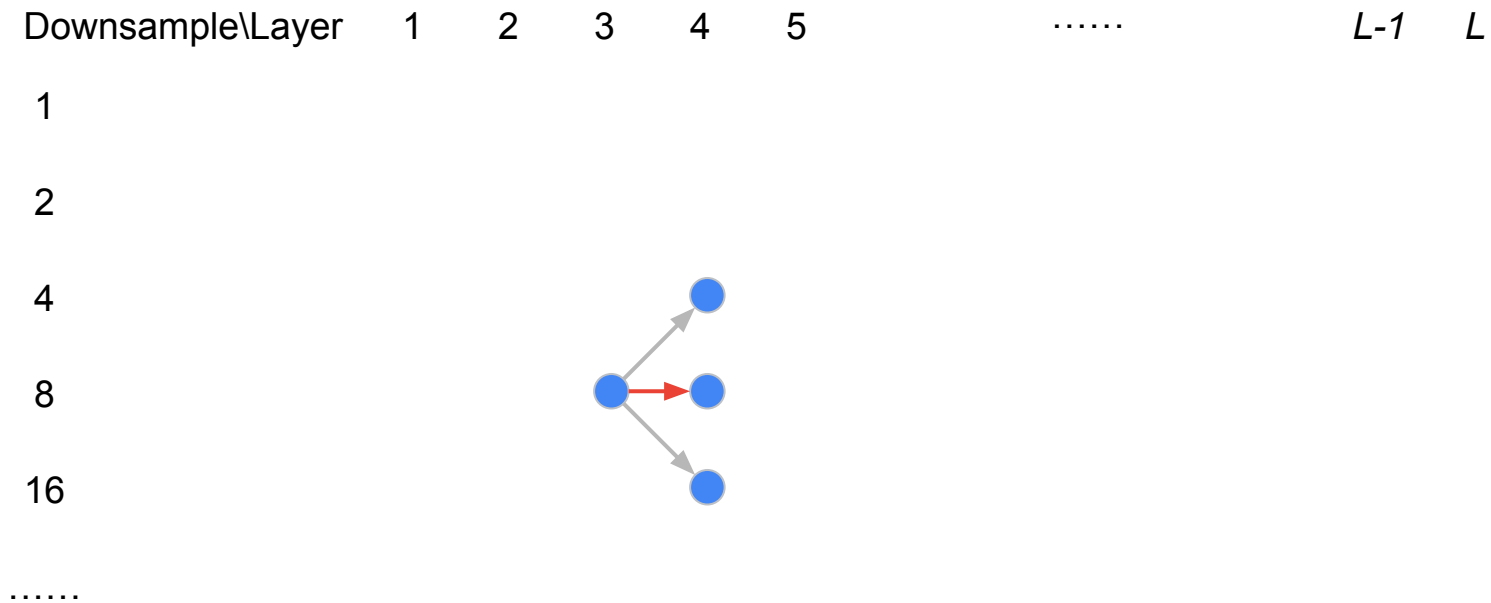
32

Network Level Search Space

DeepLabv3

Chen, Liang-Chieh, George Papandreou, Florian Schroff, and Hartwig Adam. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587 (2017).

Conv-Deconv

| Downsample\Layer | 1 | 2 | 3 | 4 | 5 | ...... | L-1 | L |

Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. "Learning deconvolution network for semantic segmentation." In ICCV. 2015.

Stacked Hourglass

Newell, Alejandro, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation." In ECCV. 2016.
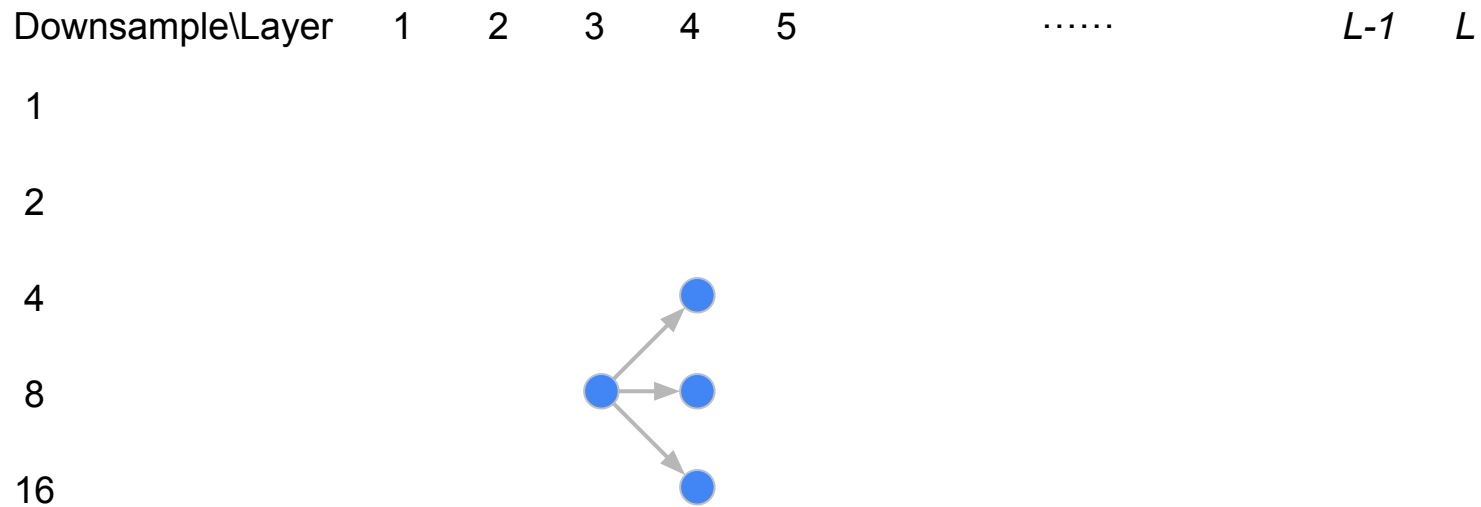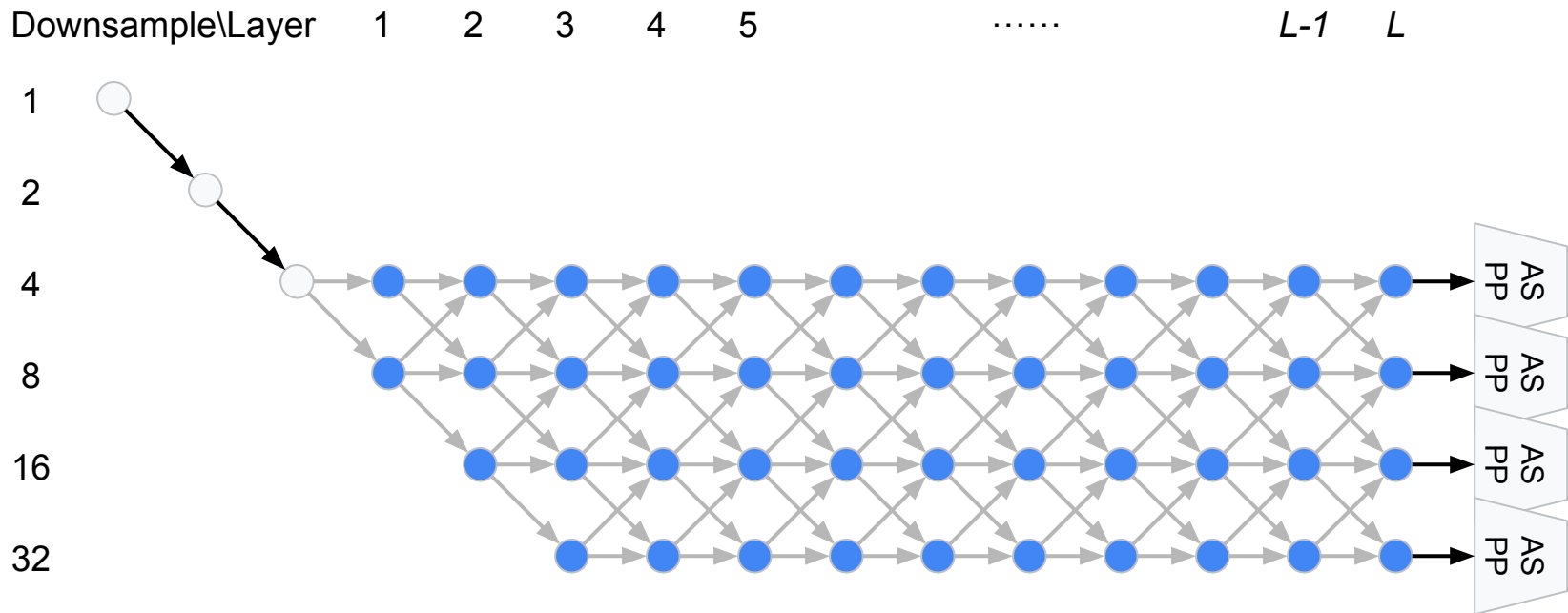
Network Level Search Space
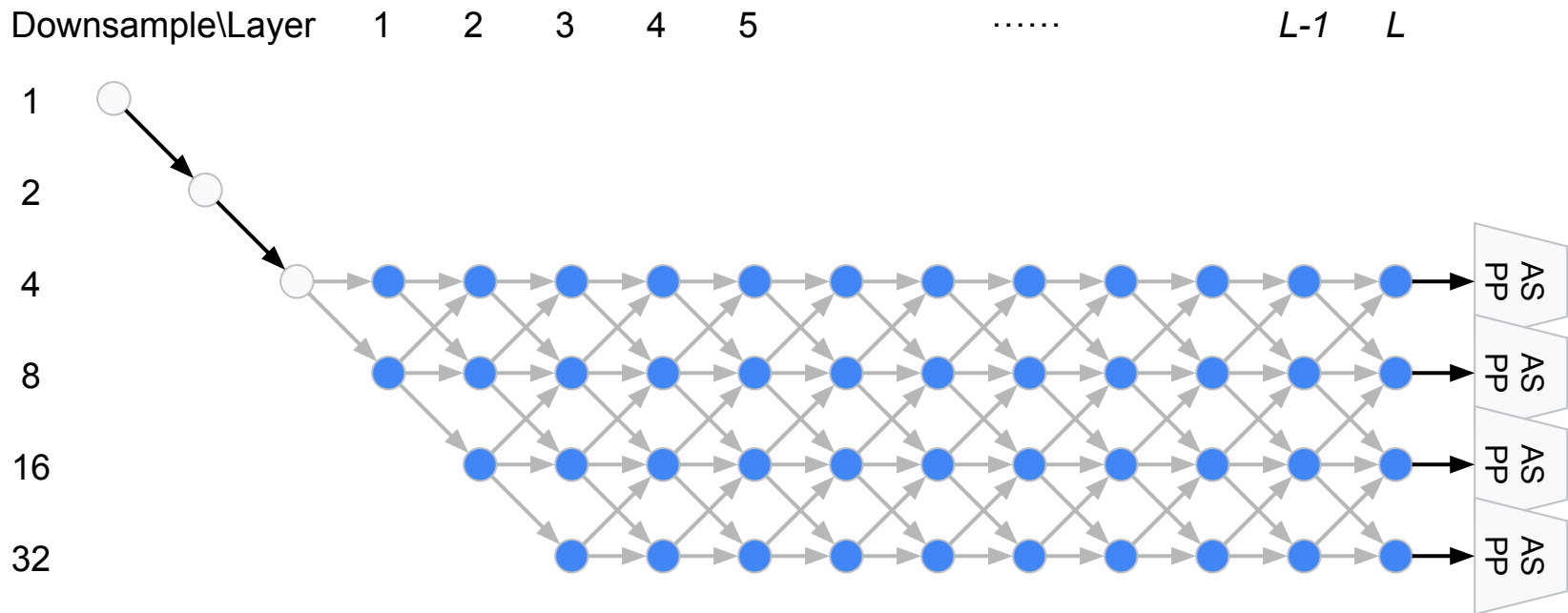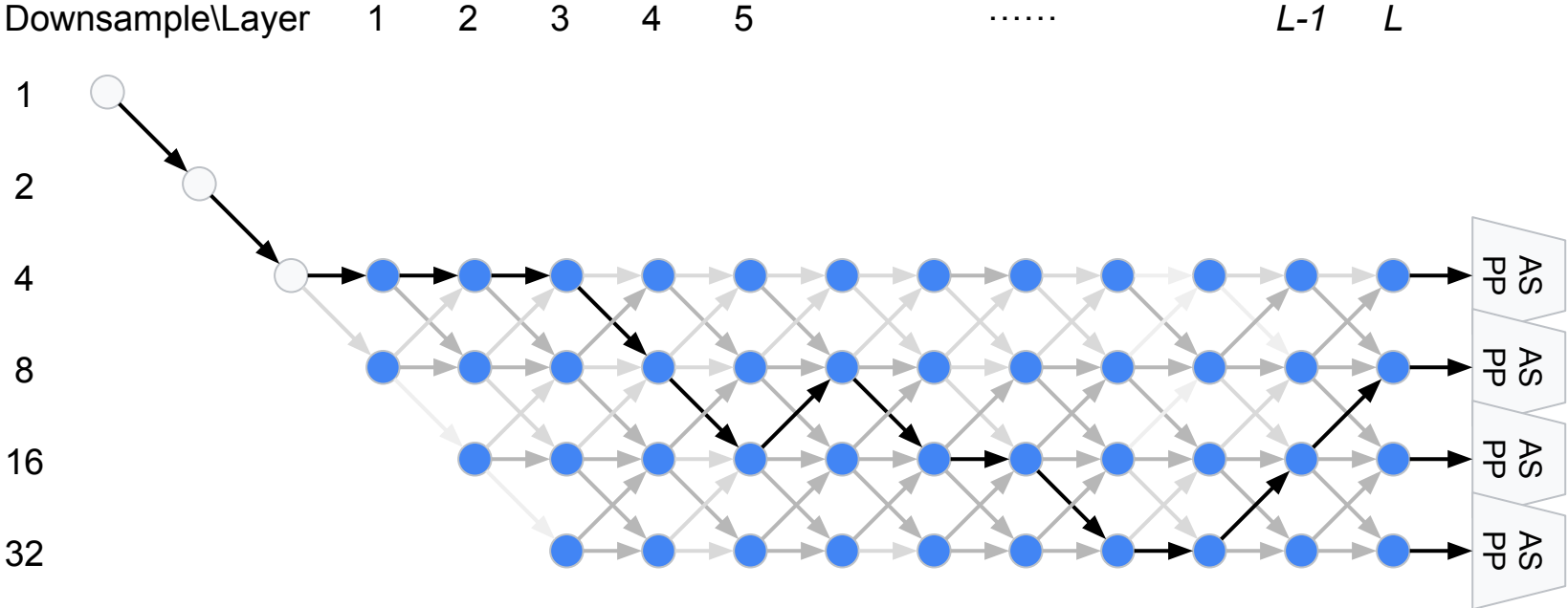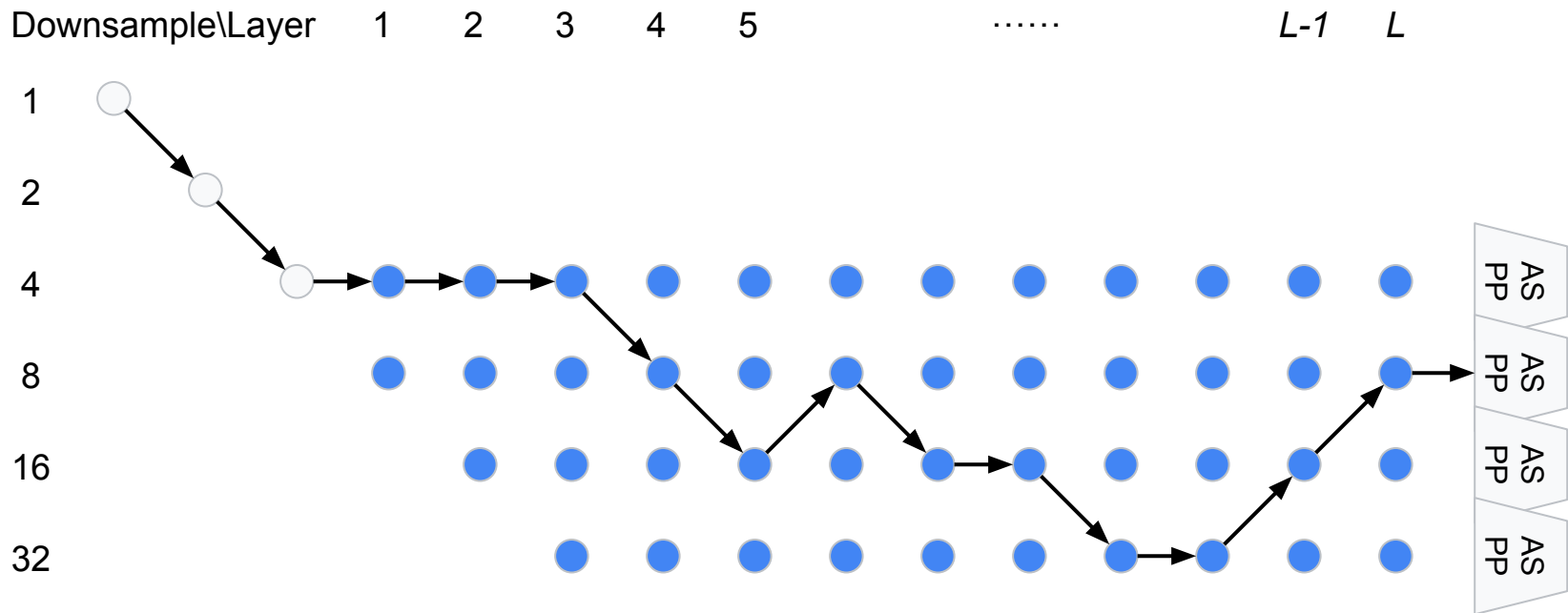
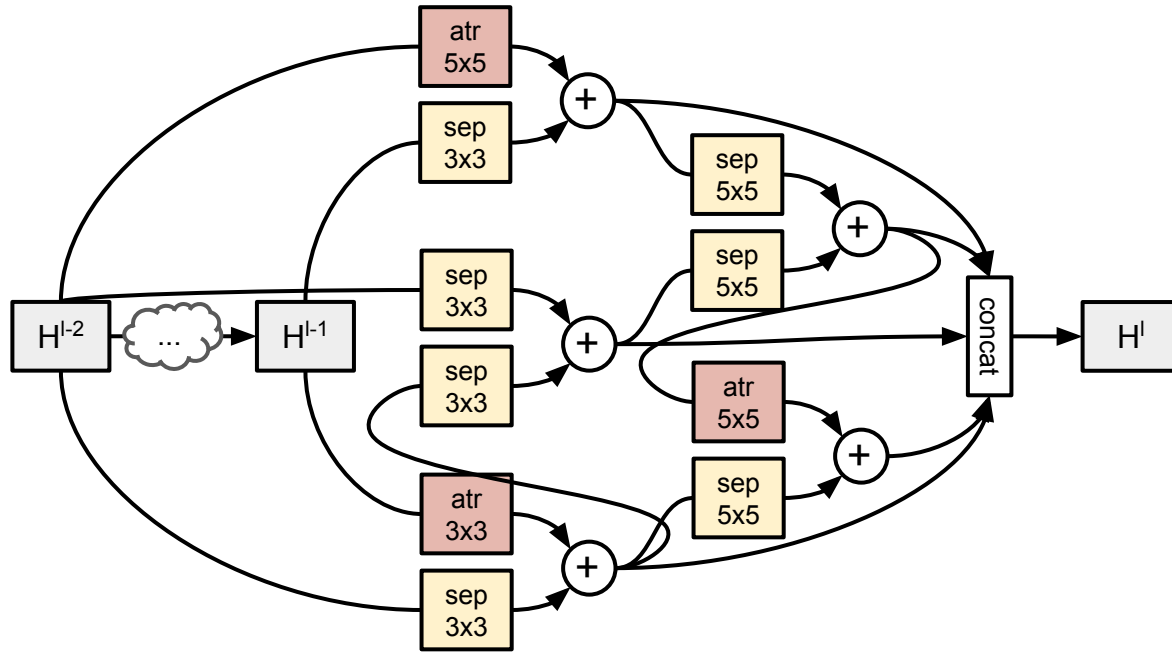# Network Level Search Space

Network Level Search Space

Experiments

- 321x321 image crops from Cityscapes

- Number of layers L = 12

- 40 epochs; **less than 3 days on one P100 GPU**

Auto-DeepLab Cell Architecture

Auto-DeepLab Cell Architecture
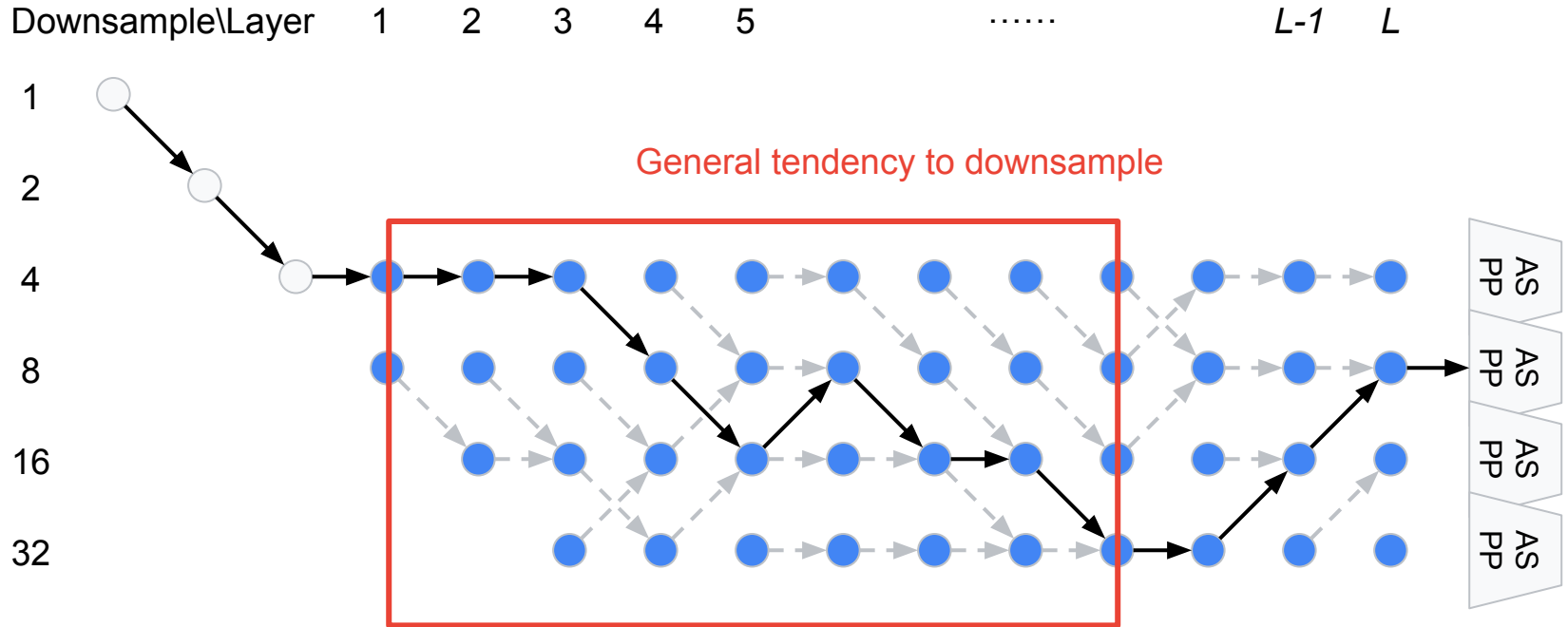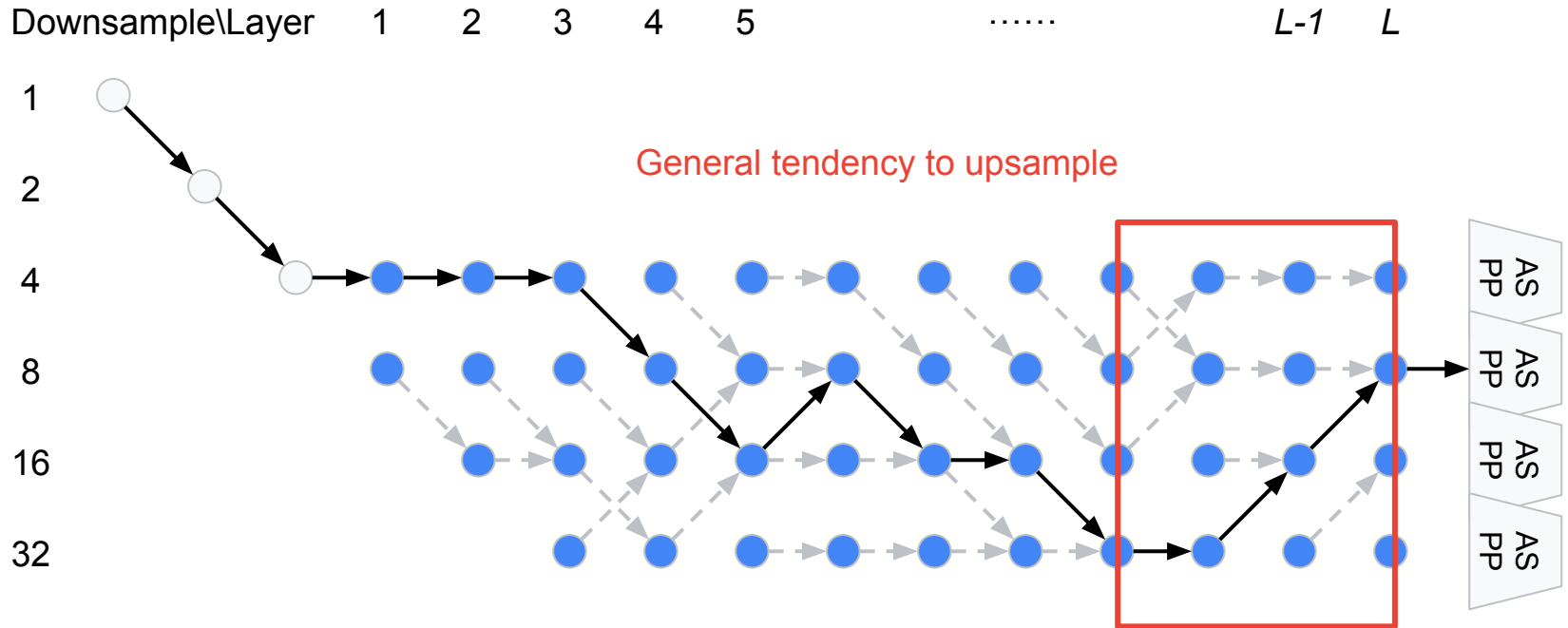

Atrous convolution is often used

Auto-DeepLab Network Architecture

Auto-DeepLab Network Architecture

Auto-DeepLab Network Architecture

Downsample\Layer   1   2   3   4   5   ……   L-1   L

General tendency to upsample

Performance on Cityscapes (Test Set)

Method	ImageNet?	Coarse?	mIOU (%)
GridNet			69.5
FRRN-B			**71.8**
Auto-DeepLab-S			79.9
Auto-DeepLab-L			**80.4**
Auto-DeepLab-S		Yes	80.9
Auto-DeepLab-L		Yes	82.1
DeepLabv3+	Yes	Yes	82.1
DPC	Yes	Yes	82.7

Fourure, Damien, et al. "Residual conv-deconv grid network for semantic segmentation." In BMVC. 2017.
Pohlen, Tobias, et al. "Full-resolution residual networks for semantic segmentation in street scenes." In CVPR. 2017.
Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." In ECCV. 2018.
Chen, Liang-Chieh, et al. "Searching for efficient multi-scale architectures for dense image prediction." In NeurIPS. 2018.

Performance on Cityscapes (Test Set)

Method	ImageNet?	Coarse?	mIOU (%)
GridNet			69.5
FRRN-B			71.8
Auto-DeepLab-S			79.9
Auto-DeepLab-L			80.4
Auto-DeepLab-S		Yes	80.9
Auto-DeepLab-L		**Yes**	**82.1**
DeepLabv3+	**Yes**	**Yes**	**82.1**
DPC	Yes	Yes	82.7

Fourure, Damien, et al. "Residual conv-deconv grid network for semantic segmentation." In BMVC. 2017.
Pohlen, Tobias, et al. "Full-resolution residual networks for semantic segmentation in street scenes." In CVPR. 2017.
Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." In ECCV. 2018.
Chen, Liang-Chieh, et al. "Searching for efficient multi-scale architectures for dense image prediction." In NeurIPS. 2018.

Thank You

@chenxi116   https://cs.jhu.edu/~cxliu/