REGION MUTUAL INFORMATION LOSS FOR SEMANTIC SEGMENTATION

Shuai Zhao[1], Yang Wang[2], Zheng Yang[3], Deng Cai[1]
Author Affiliations: 1 - Zhejiang University, 2 - Huazhong University and 3 - Fabu Inc.

NeurIPS 2019

Presented by: Rohit Gupta

tl;dr / one line summary

Novel loss function (RMI)

+

Existing segmentation methods (DeepLabv3/DeepLabv3+)

=

SotA Results

Sections

- Semantic Segmentation
- Overview of Older Methods: FCN, U-Net, DeepLab, DeepLabv2
- DeepLabv3 & DeepLabv3+
- Mutual Information & RMI Loss
- Results and Ablations
- Discussion

Semantic Segmentation

Input



Prediction

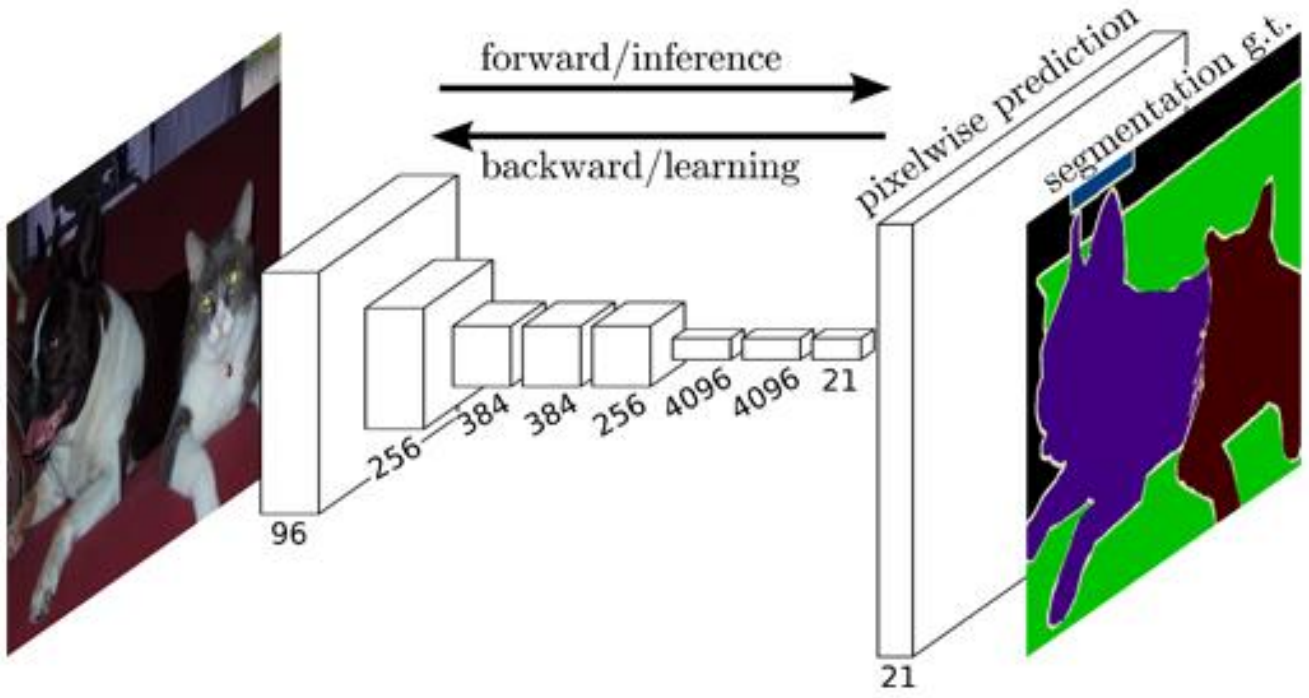


Images and annotations from MS-COCO copyright rests with original owners or authors.

FCN: Fully Convolutional Network

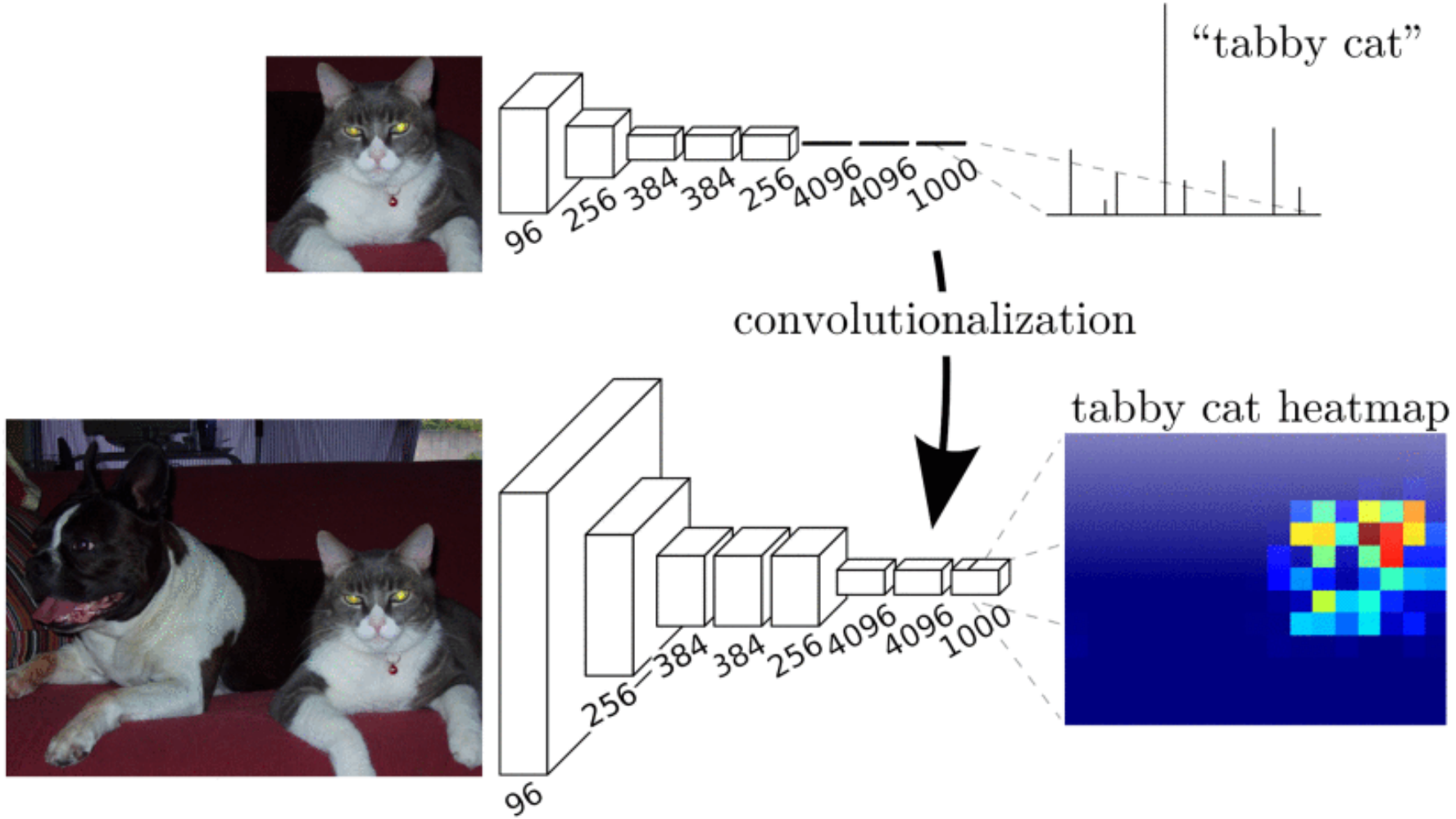
Jonathan Long*, Evan Shelhamer* & Trevor Darrell

CVPR 2015 & TPAMI Vol. 39 Issue 4 (Apr 2017) **13,462 citations**



All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea: “convolutional-ization”

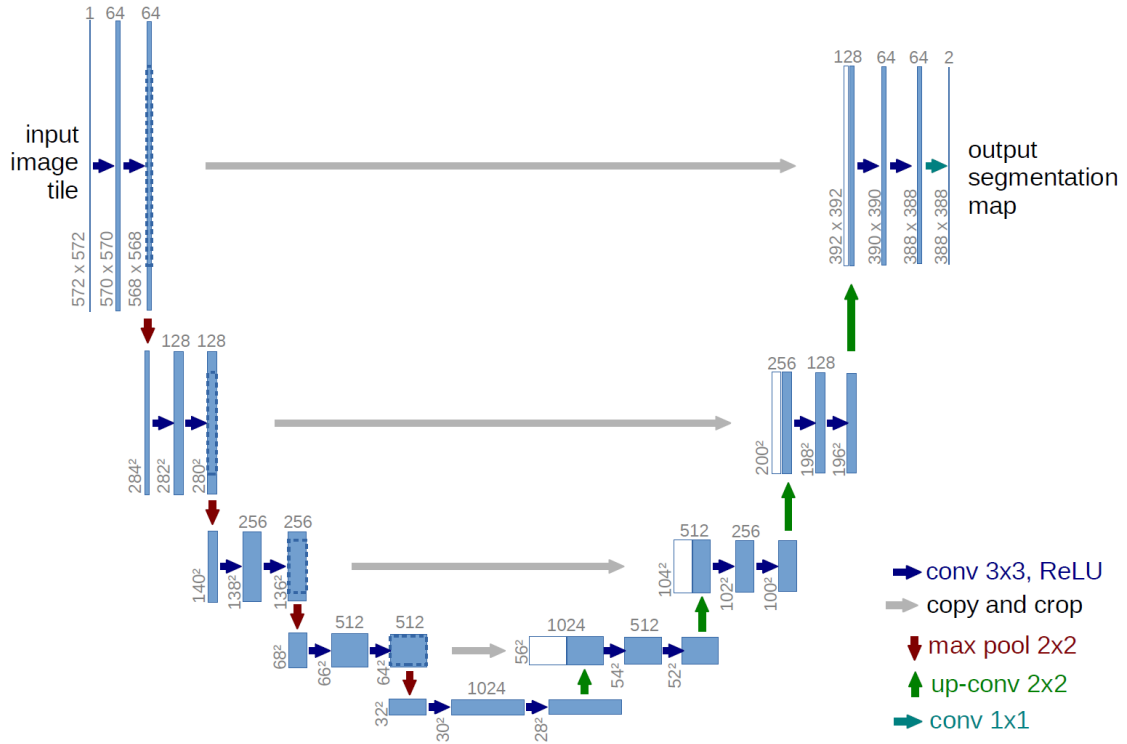


All figures and other images are from the respective papers, copyright rests with original owners or authors.

U-Net : Convolutional Networks for Biomedical Image Segmentation

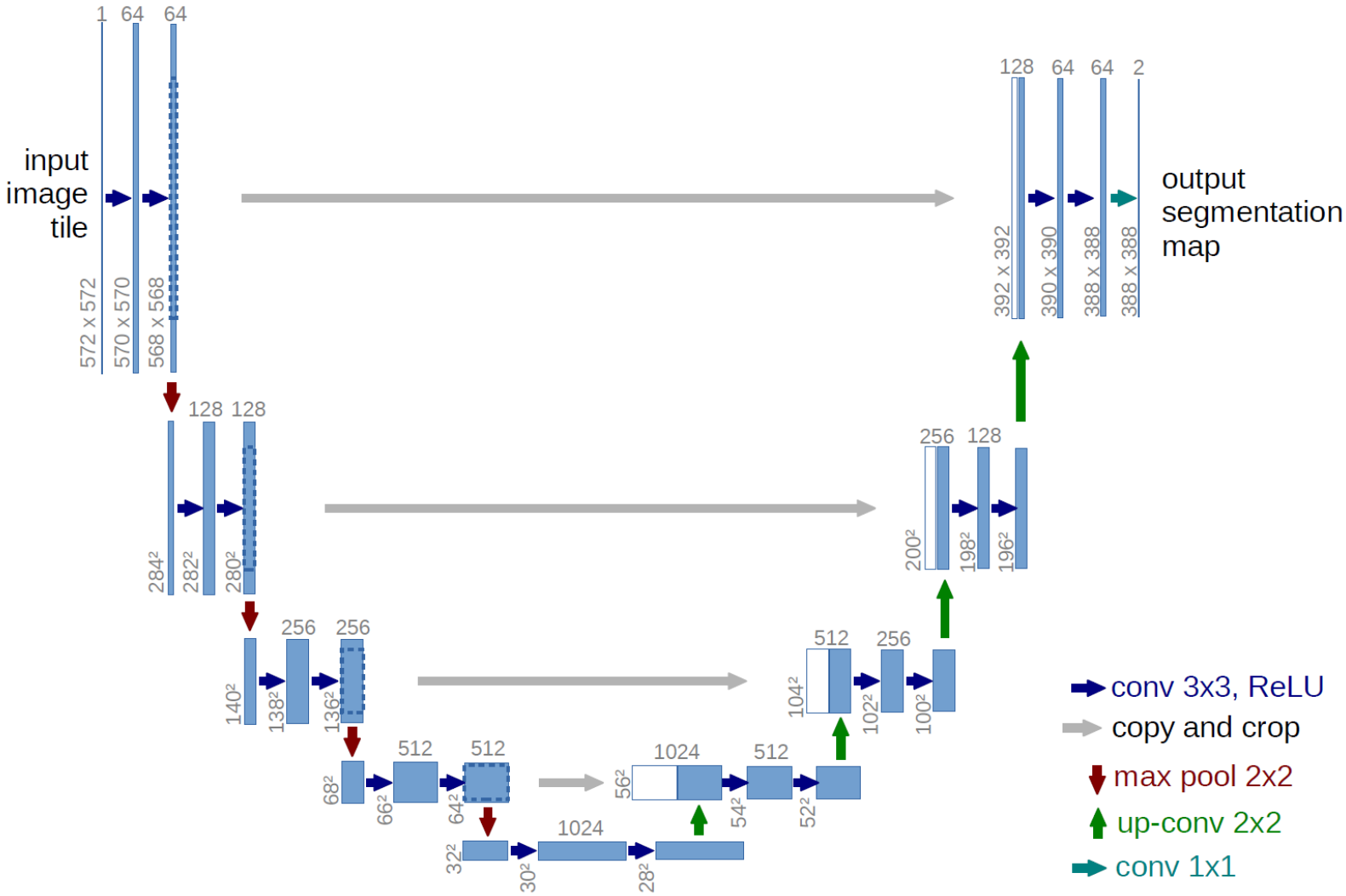
Olaf Ronneberger, Philipp Fischer, Thomas Brox

MICCAI 2015 **10,945 Citations**



All figures and other images are from the respective papers, copyright rests with original owners or authors.

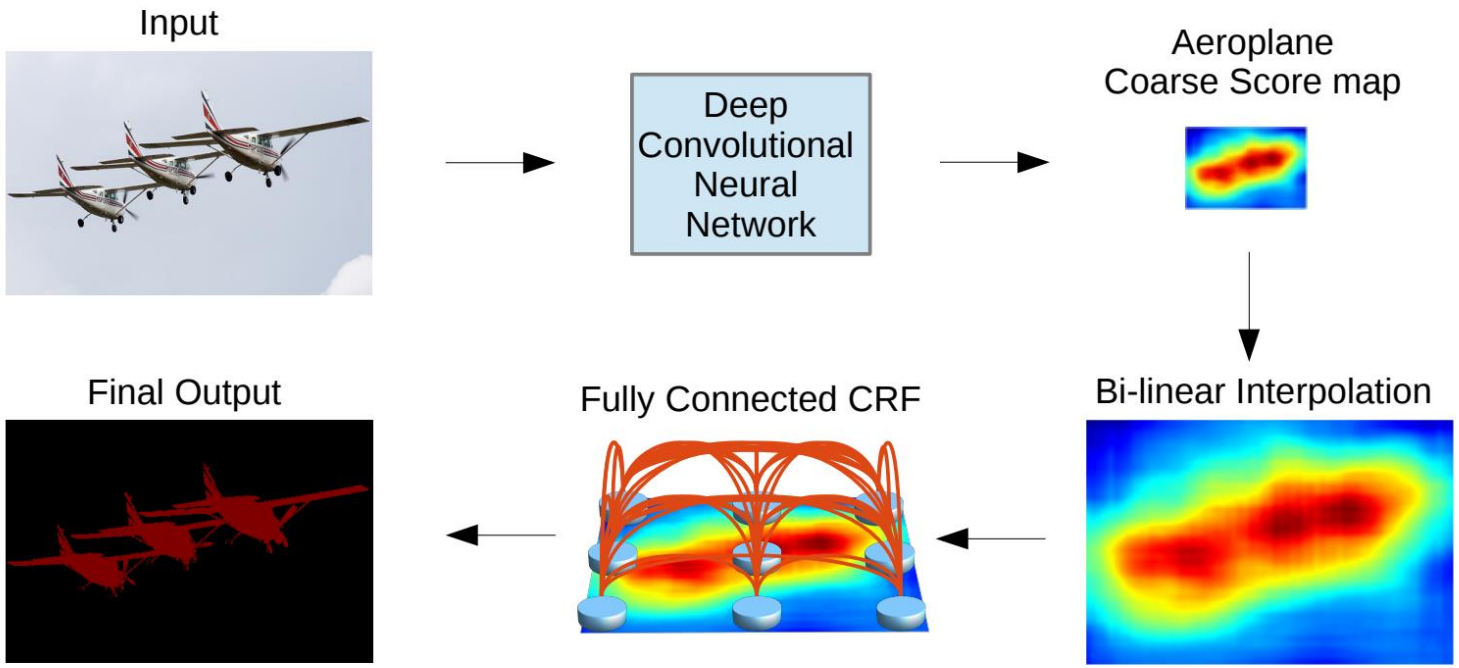
Key Idea: Encoder-decoder architecture



All figures and other images are from the respective papers, copyright rests with original owners or authors.

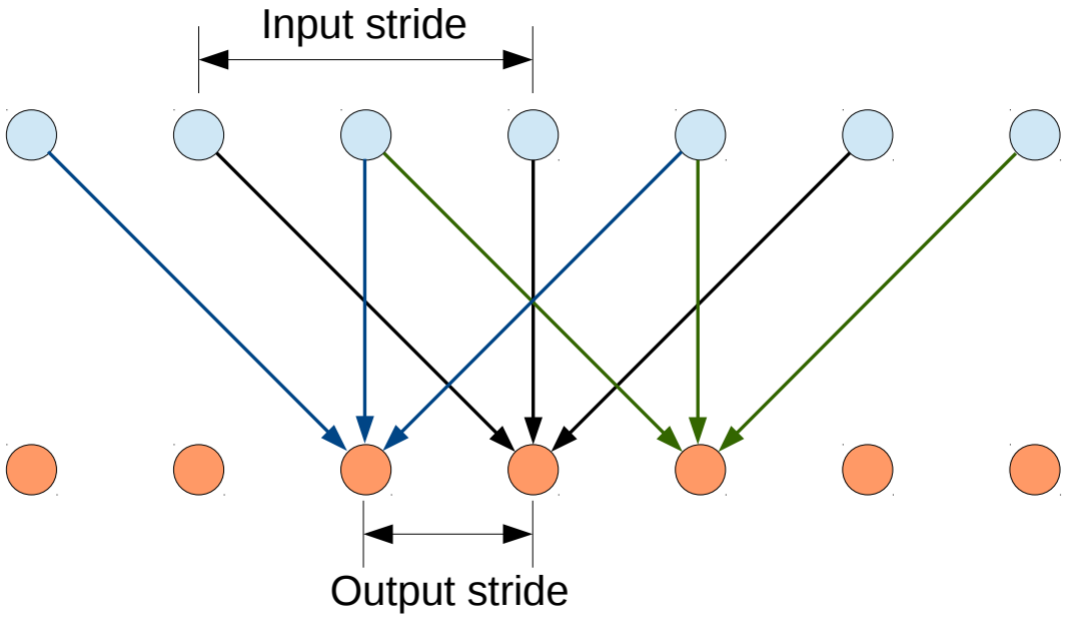
DeepLab : Semantic Image Segmentation w/ CNNs & Fully Connected CRFs

Liang-Chieh Chen*, George Papandreou* & others (UCLA, Google & INRIA)
ICLR 2015 **2,138 Citations**



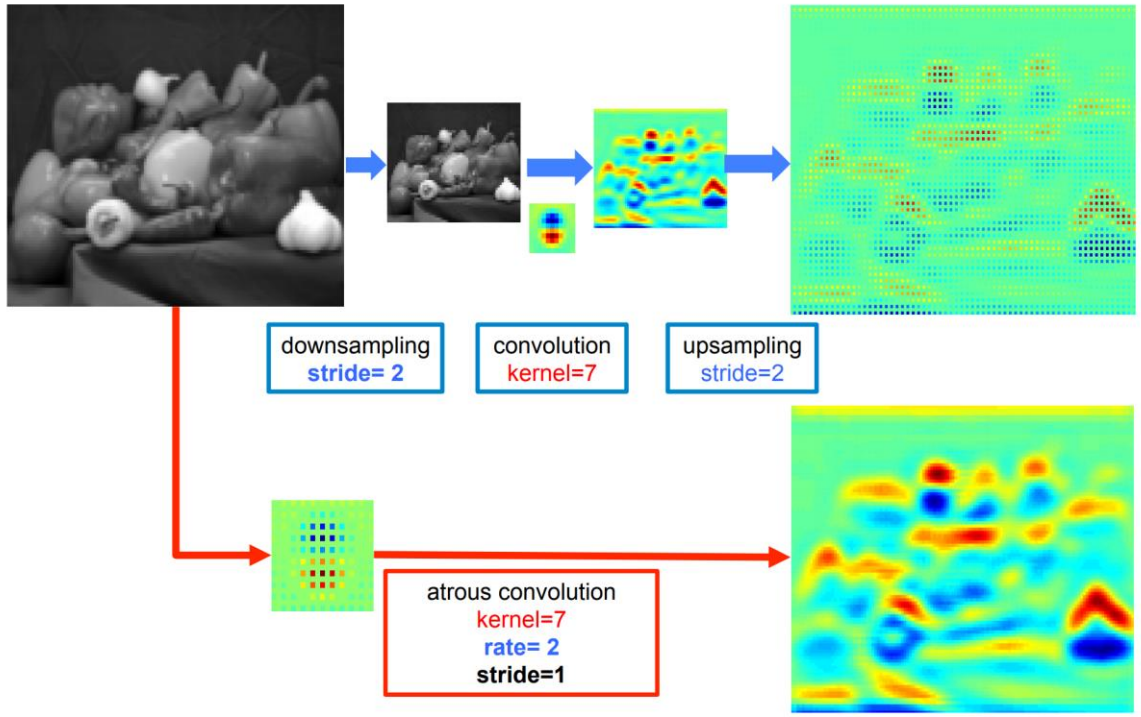
All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea #1: Convolutionalization with “à trous convolutions”



All figures and other images are from the respective papers, copyright rests with original owners or authors.

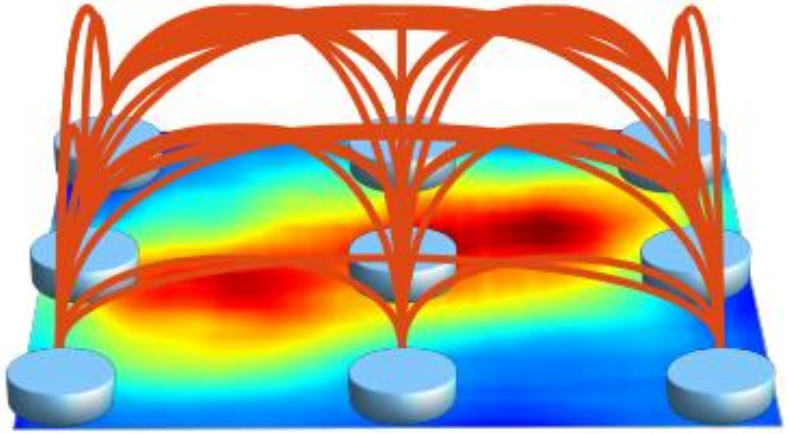
Key Idea #1: Convolutionalization with “à trous convolutions”



All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea #2: CRF Post-Processing

Fully Connected CRF

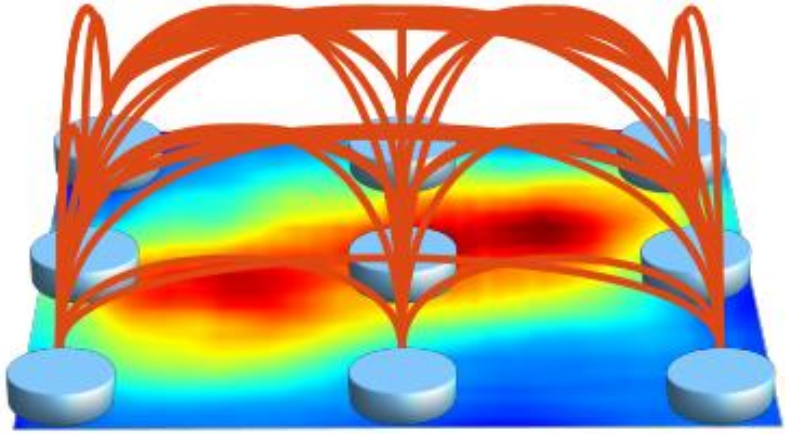


All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea #2: CRF Post-Processing

$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

Fully Connected CRF



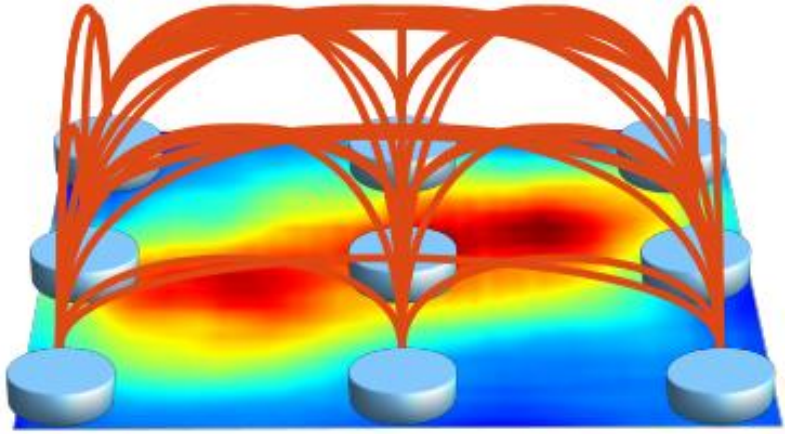
All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea #2: CRF Post-Processing

$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

$$\theta_i(x_i) = -\log P(x_i)$$

Fully Connected CRF



All figures and other images are from the respective papers, copyright rests with original owners or authors.

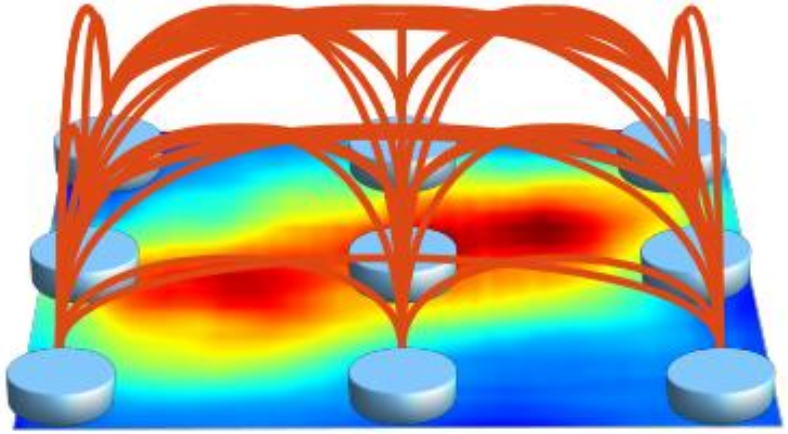
Key Idea #2: CRF Post-Processing

$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

$$\theta_i(x_i) = -\log P(x_i)$$

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w_m \cdot k^m(\boldsymbol{f}_i, \boldsymbol{f}_j)$$

Fully Connected CRF

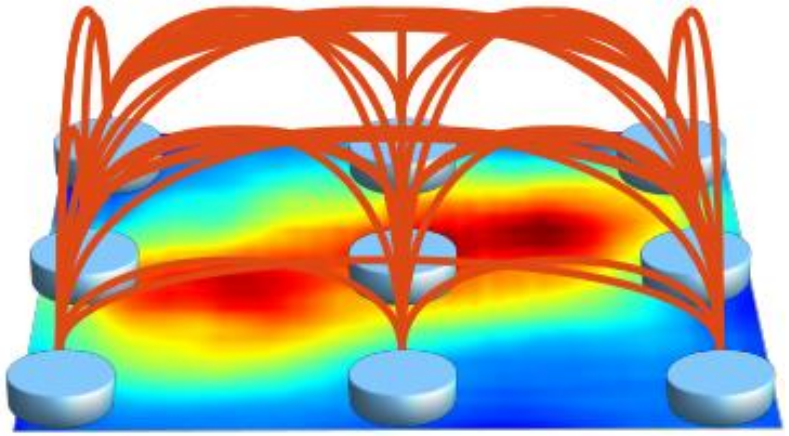


All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea #2: CRF Post-Processing

Fully Connected CRF

$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$



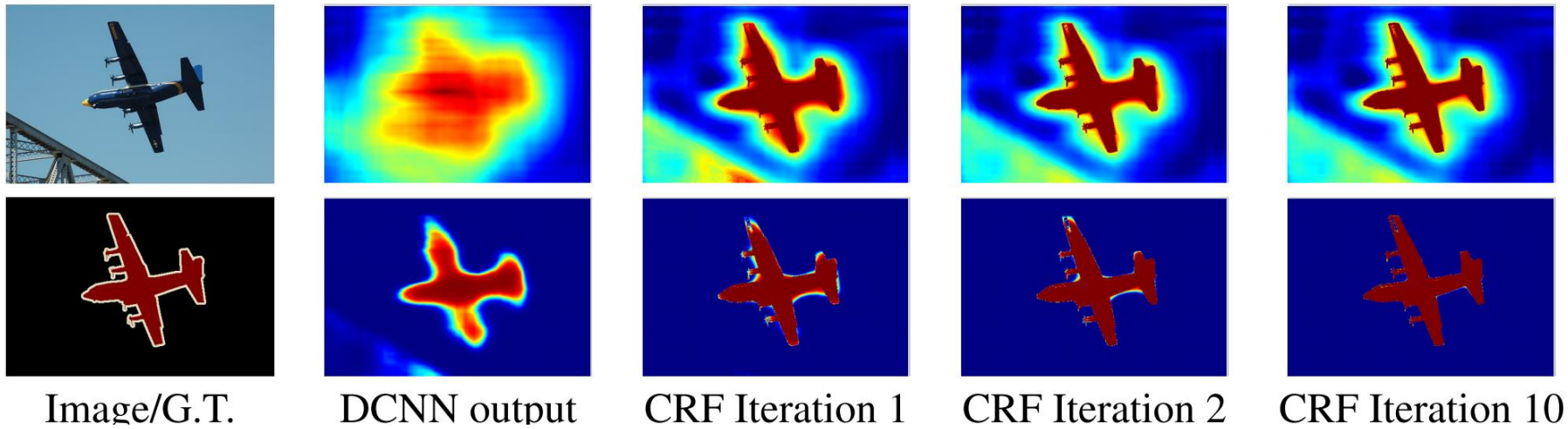
$$\theta_i(x_i) = -\log P(x_i)$$

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w_m \cdot k^m(\boldsymbol{f}_i, \boldsymbol{f}_j)$$

$$w_1 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\alpha^2} - \frac{||I_i - I_j||^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\gamma^2}\right)$$

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea #2: CRF Post-Processing

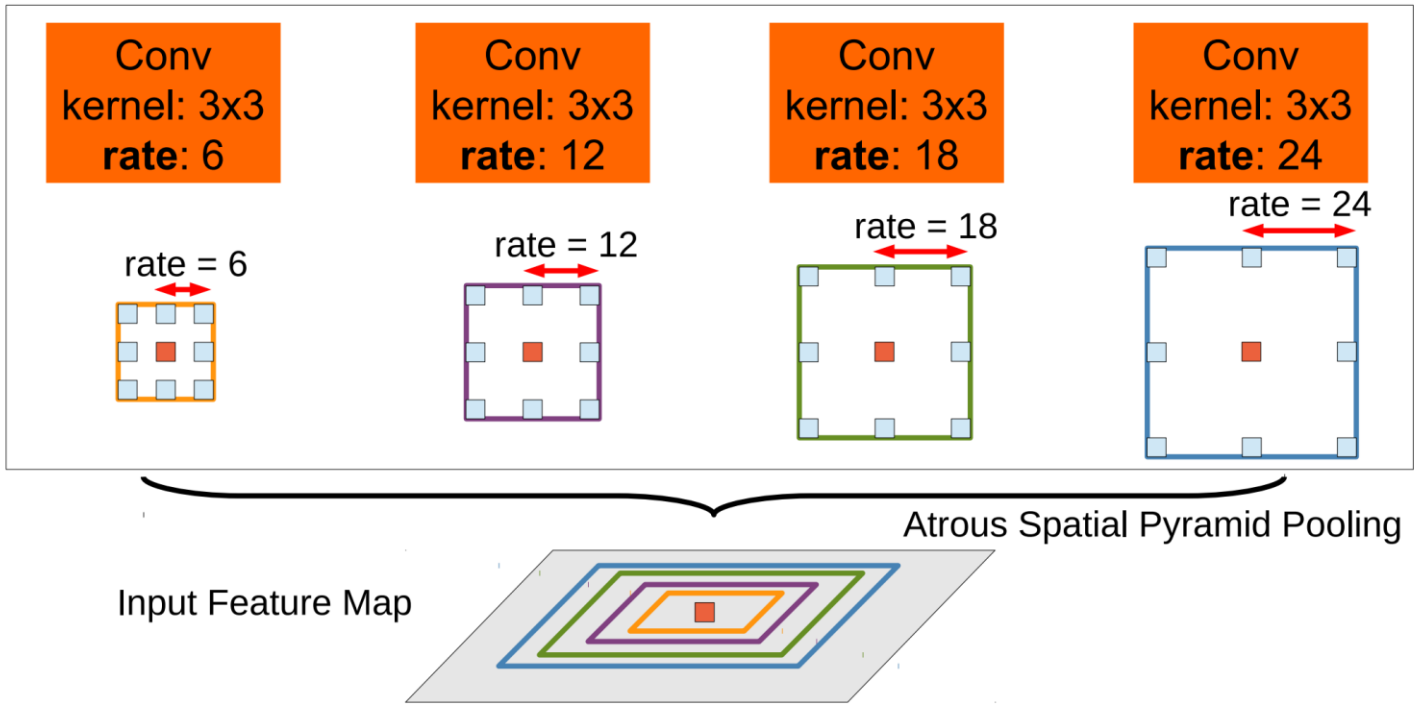


All figures and other images are from the respective papers, copyright rests with original owners or authors.

DeepLabv2 : Semantic Image Segmentation w/ CNNs, Atrous Convolution, & Fully Connected CRFs

Liang-Chieh Chen*, George Papandreou* & others (Google ?)

TPAMI Vol. 40 Issue 4 (2018) **3,918 Citations**



All figures and other images are from the respective papers, copyright rests with original owners or authors.

DeepLabv3 : Rethinking Atrous Convolution for Semantic Image Segmentation

Liang-Chieh Chen, George Papandreou & others (Google)
arXiv:1706.05587 **1,052 Citations**

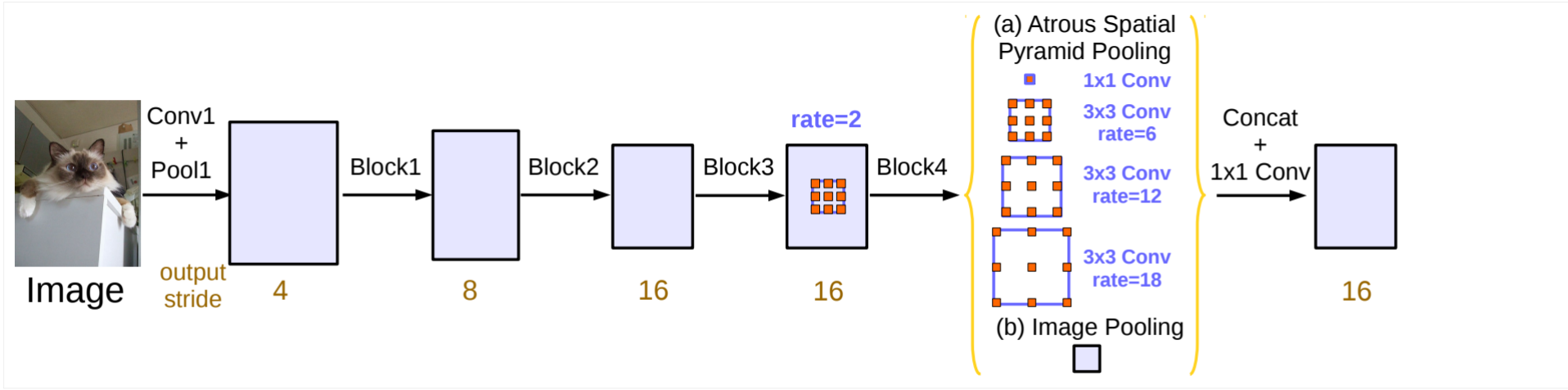
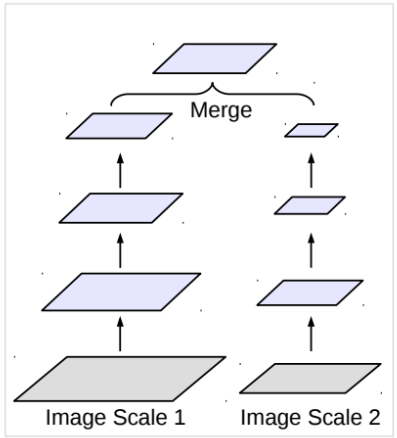


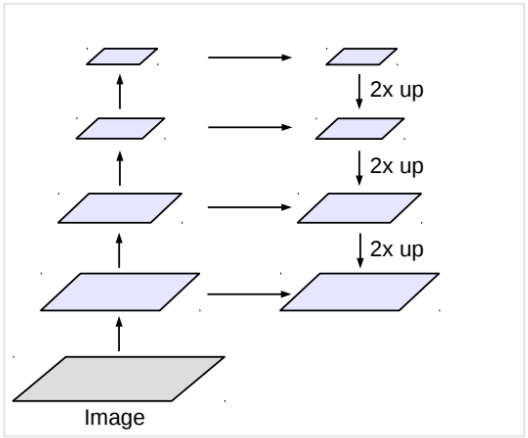
Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

All figures and other images are from the respective papers, copyright rests with original owners or authors.

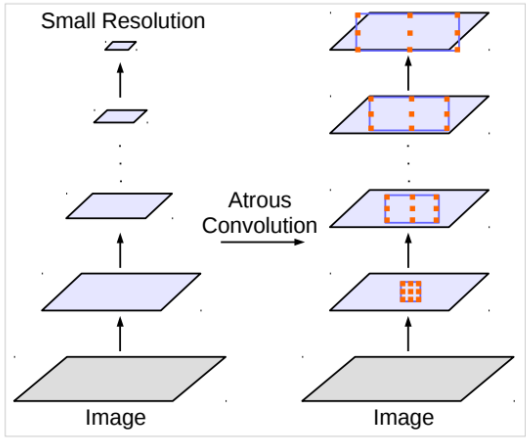
Key Idea: Multi-Scale Features



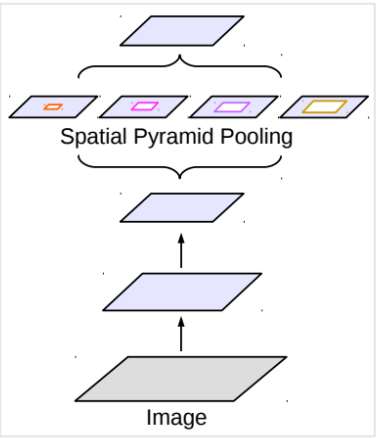
(a) Image Pyramid



(b) Encoder-Decoder



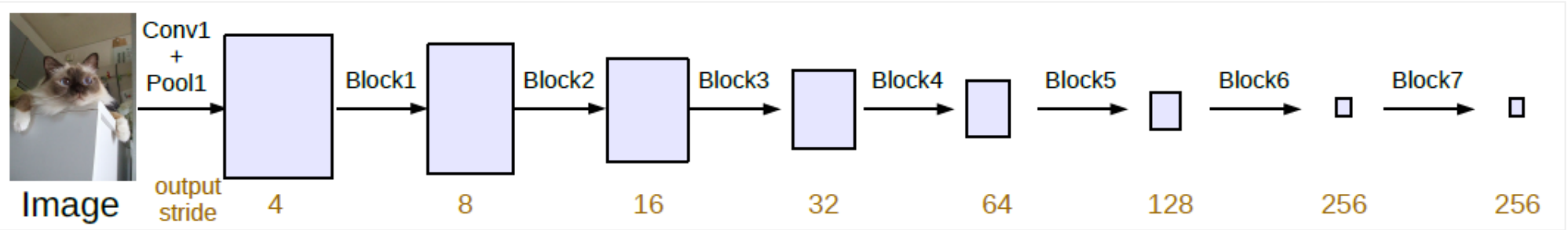
(c) Deeper w. Atrous Convolution



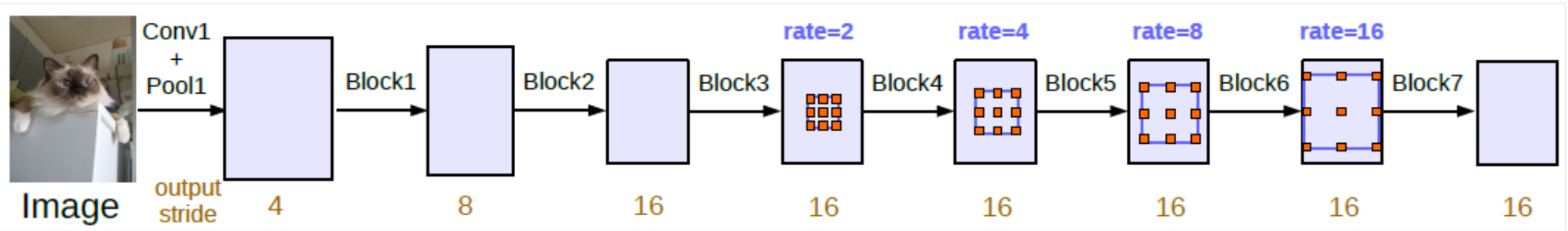
(d) Spatial Pyramid Pooling

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea: Deeper with à trous convolutions



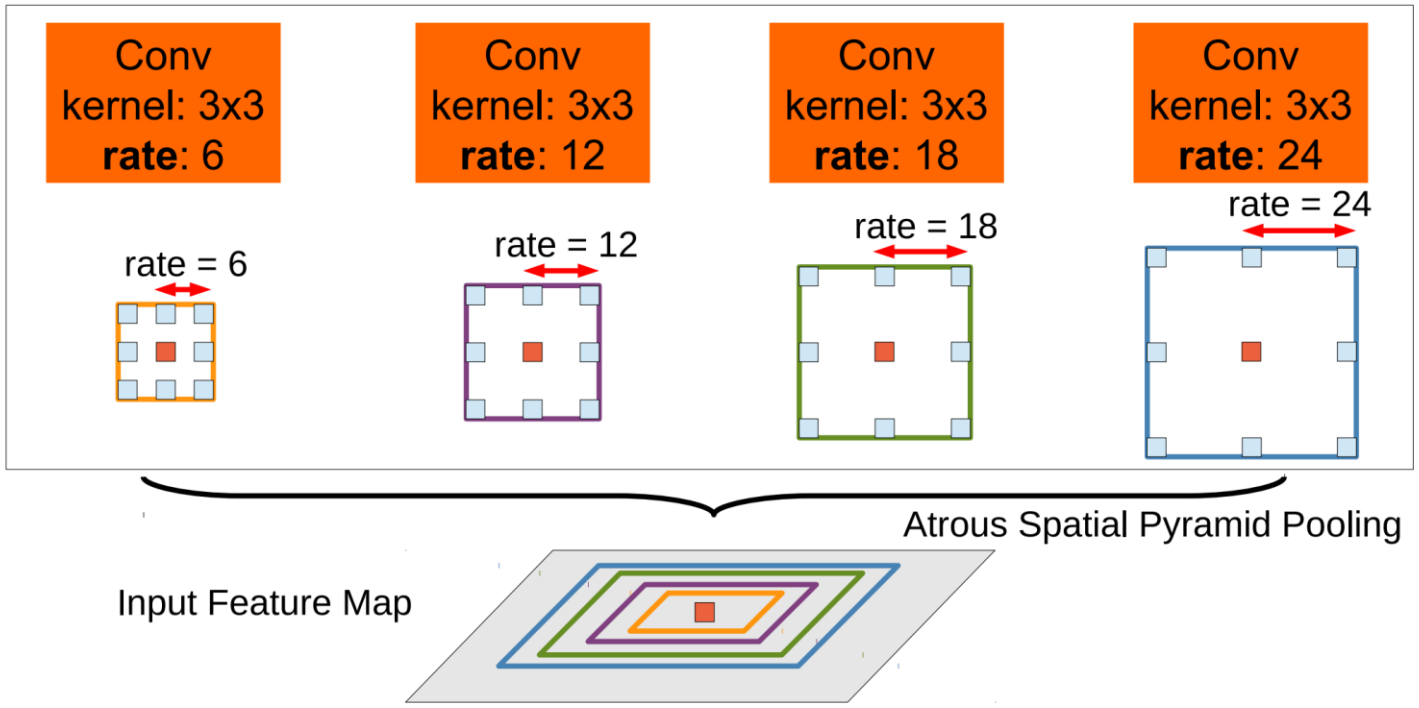
(a) Going deeper without atrous convolution.



(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output_stride = 16$.

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea: **À** trous **S**patial **P**yramid **P**ooling

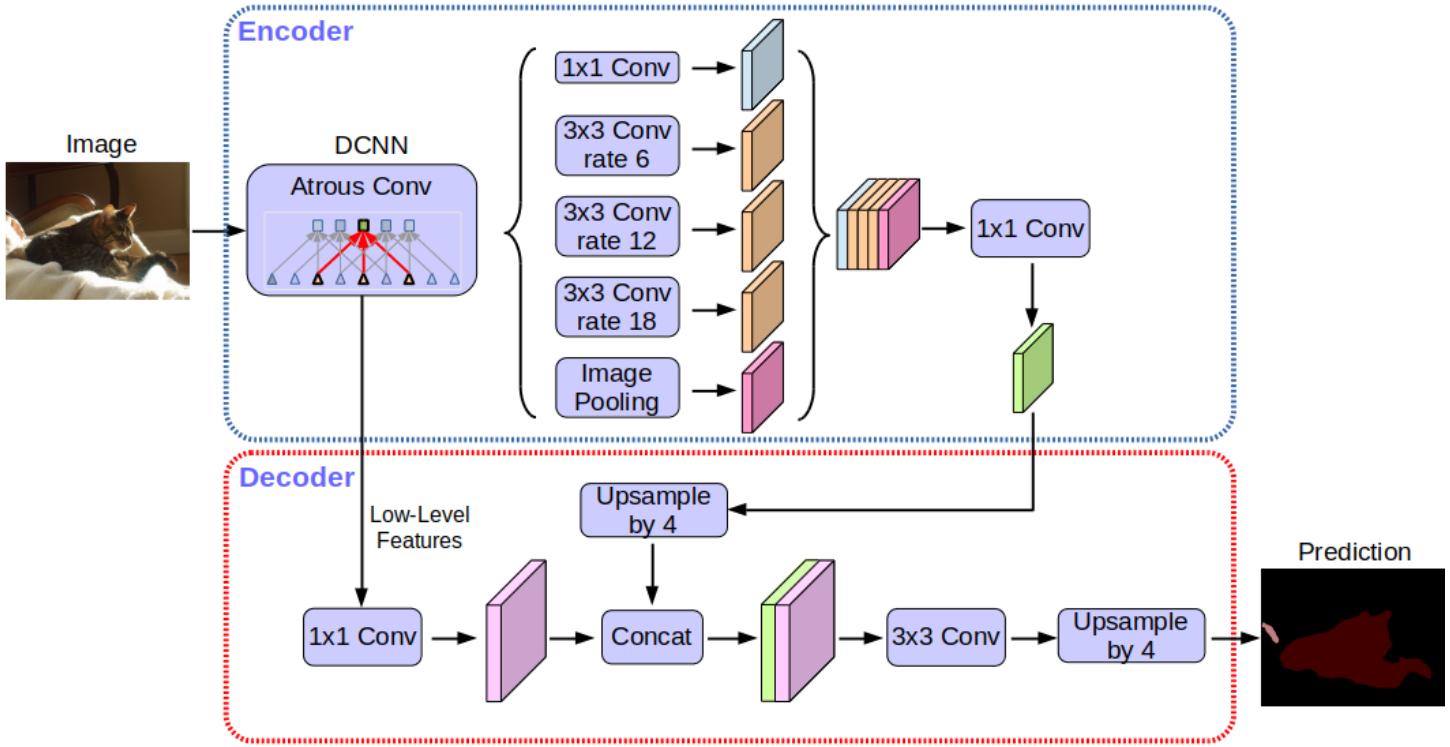


All figures and other images are from the respective papers, copyright rests with original owners or authors.

DeepLabv3+ : Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation

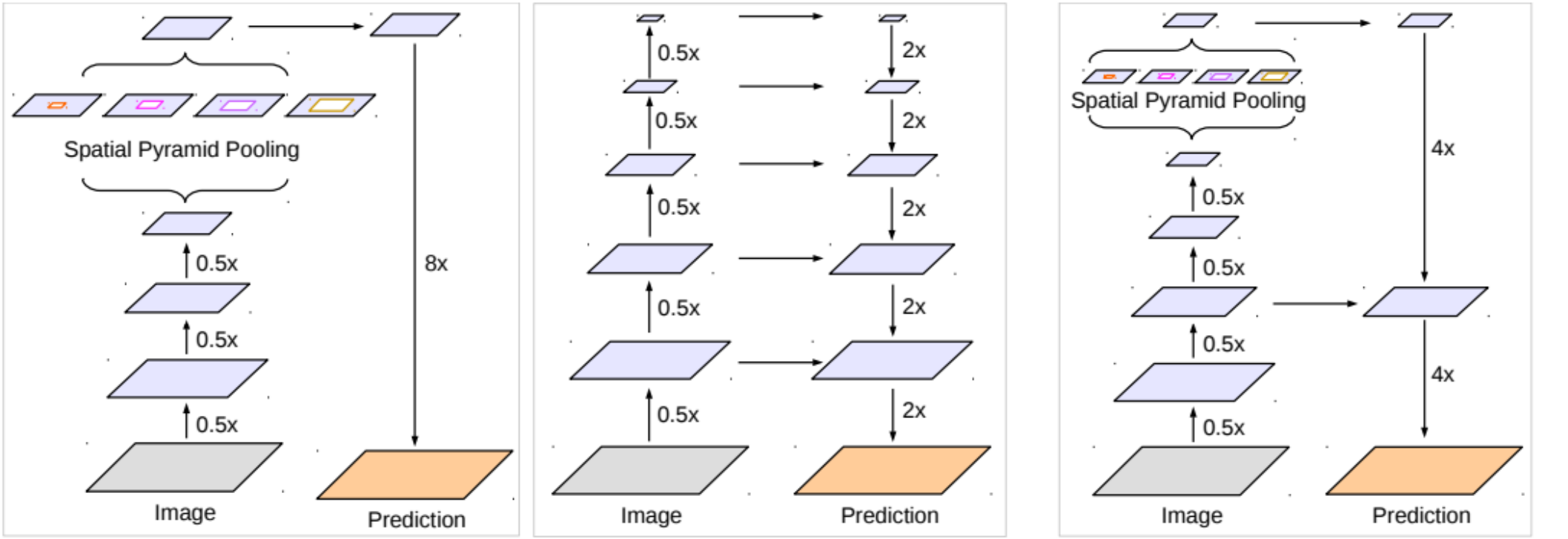
Liang-Chieh Chen & others (Google)

ECCV 2018 **1,107 Citations**



All figures and other images are from the respective papers, copyright rests with original owners or authors.

Key Idea: Encoder-Decoder + ASPP

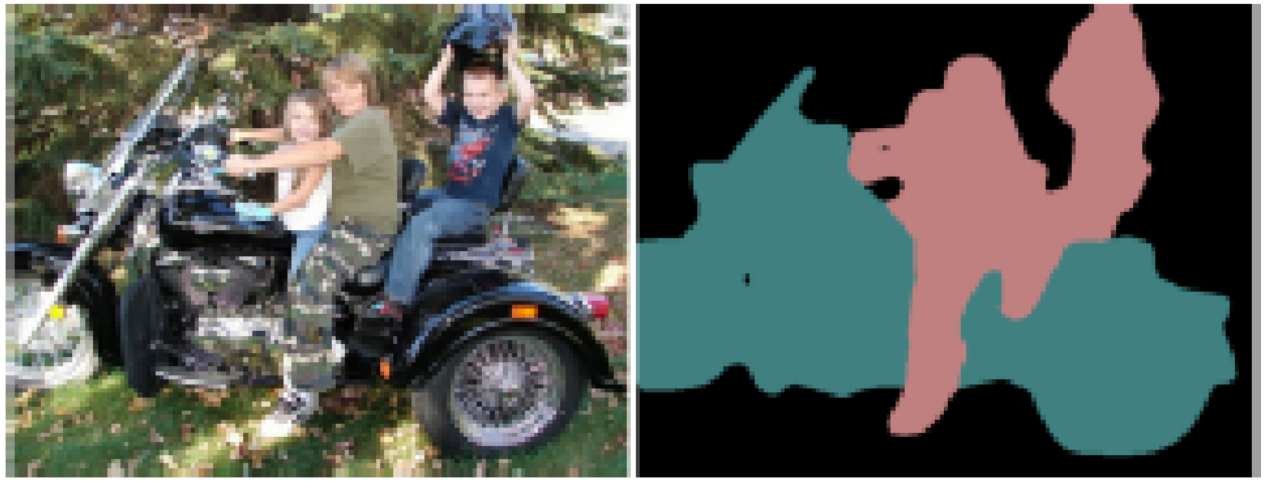


(a) Spatial Pyramid Pooling (b) Encoder-Decoder (c) Encoder-Decoder with Atrous Conv

All figures and other images are from the respective papers, copyright rests with original owners or authors.

DeepLabv3+: Qualitative Results





All figures and other images are from the respective papers, copyright rests with original owners or authors.

DeepLabv3+: Failure Cases



All figures and other images are from the respective papers, copyright rests with original owners or authors.

DeepLabv3+: Failure Cases





All figures and other images are from the respective papers, copyright rests with original owners or authors.

Motivation for Region Mutual Information Loss

- Spatial smoothness/coherence of predictions
- Pixelwise loss function

Motivation for Region Mutual Information Loss

- Spatial smoothness/coherence of predictions
- Pixelwise loss function
- CRF not used in DeepLabv3 and DeepLabv3+
  - Not real time: ~1 second per image
- Fast: shouldn’t increase inference time

Mutual Information

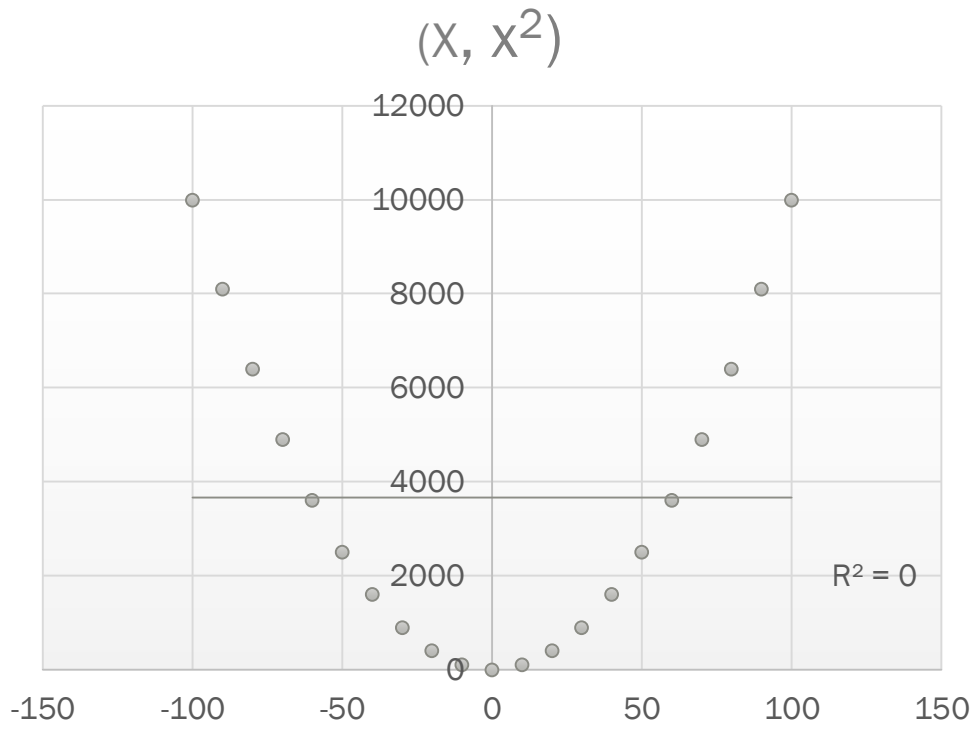
- Quantifies “amount of information” obtained about one random variable through observing the other random variable

Mutual Information

- Quantifies “amount of information” obtained about one random variable through observing the other random variable
- Unlike correlation it is not limited to capturing linear dependence
- E.g. Correlation between x and x^2 is 0

Mutual Information

- Quantifies “amount of information” obtained about one random variable through observing the other random variable
- Unlike correlation it is not limited to capturing linear dependence
- E.g. Correlation between x and x^2 is 0



Mutual Information: Entropy View

- Entropy ← "uncertainty"
- $\mathrm{H(x)} = -\sum_{x \epsilon \chi} \mathrm{p(x)}\ \log p(x)$

Mutual Information: Entropy View

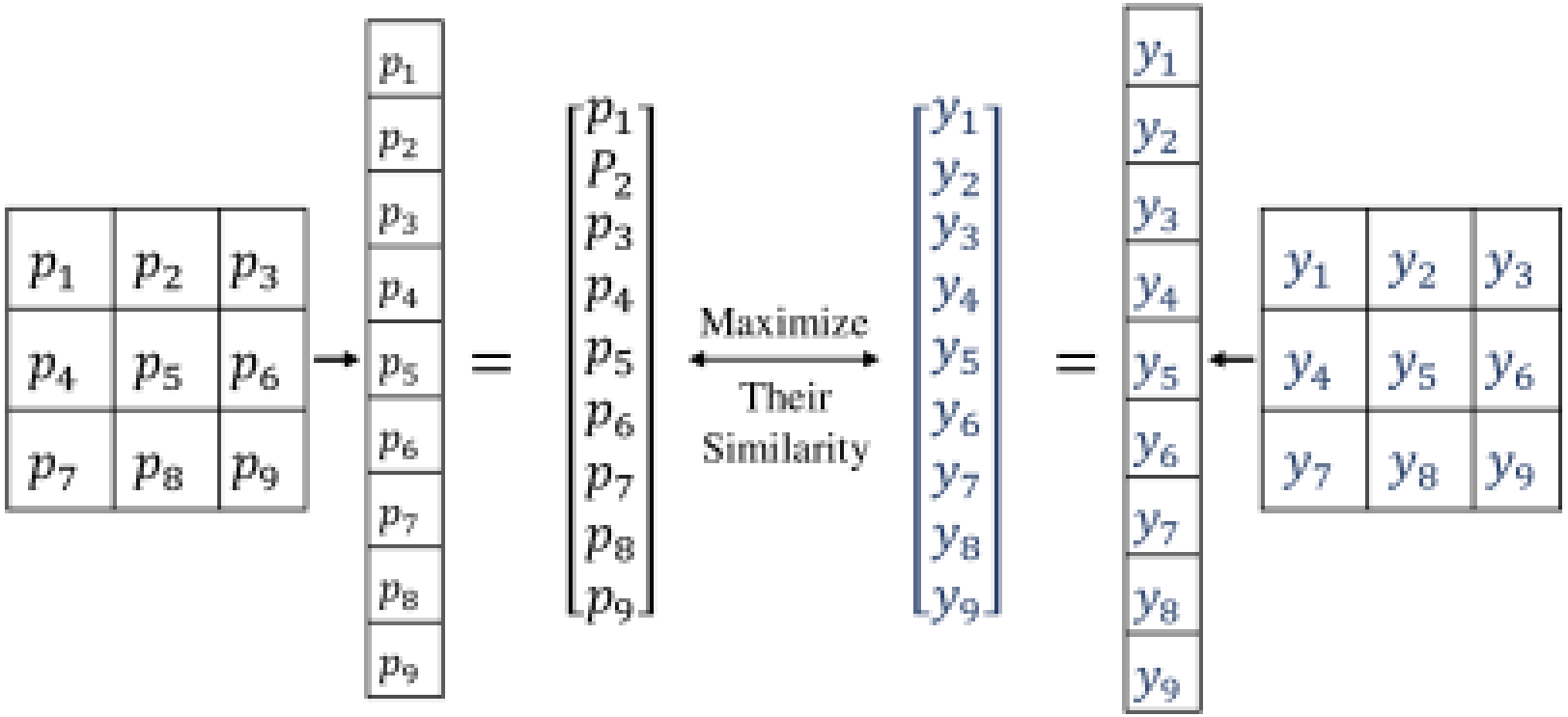
- Entropy ← "uncertainty"
- $\mathrm{H(x)} = -\sum_{x \epsilon \chi} \mathrm{p(x)}\ \log p(x)$
- Mutual Information ← "reduction in uncertainty"
- $I(X;Y) = H(X)\ -\ H(X|Y) = H(Y)\ -\ H(Y|X)$

Mutual Information: Probability View

- Entropy ← "uncertainty"
- $\mathrm{H(x)} = -\sum_{x \epsilon \chi} \mathrm{p(x)}\ \log p(x)$
- Mutual Information ← "reduction in uncertainty"
- $I(X;Y) = H(X)\ -\ H(X|Y) = H(Y)\ -\ H(Y|X)$

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Region Mutual Information



All figures and other images are from the respective papers, copyright rests with original owners or authors.

Overall Loss Function

$$\mathcal{L}_{all}(y,p) = \lambda \mathcal{L}_{ce}(y,p) + (1-\lambda)\frac{1}{B}\sum_{b=1}^{B}\sum_{c=1}^{C}\left(-I_l^{b,c}(\boldsymbol{Y};\boldsymbol{P})\right)$$

Overall Loss Function

$$\mathcal{L}_{all}(y,p) = \lambda \mathcal{L}_{ce}(y,p) + (1-\lambda)\frac{1}{B}\sum_{b=1}^{B}\sum_{c=1}^{C}\left(-I_l^{b,c}(\boldsymbol{Y};\boldsymbol{P})\right)$$

- $I_l(\mathrm{Y};\mathrm{P}) = -\frac{1}{2d}\log(\det(M))$

Overall Loss Function

$$\mathcal{L}_{all}(y,p) = \lambda \mathcal{L}_{ce}(y,p) + (1-\lambda)\frac{1}{B}\sum_{b=1}^{B}\sum_{c=1}^{C}\left(-I_l^{b,c}(\boldsymbol{Y};\boldsymbol{P})\right)$$

- $I_l(\mathrm{Y};\mathrm{P}) = -\frac{1}{2d}\log(\det(M))$
- $\mathrm{M} = \Sigma_Y - \mathrm{Cov}(\mathrm{Y},\mathrm{P})\Sigma_P^{-1^T}\mathrm{Cov}(\mathrm{Y},\mathrm{P})^T$
- $\mathrm{M} \approx \mathrm{Var}(Y|P=p)$

Experimental Setup

- Datasets: PASCAL VOC 2012, CamVid

Experimental Setup

- Datasets: PASCAL VOC 2012, CamVid
- Models: DeepLabv3, DeepLabv3+ with ResNet101 backbone

Experimental Setup

- Datasets: PASCAL VOC 2012, CamVid
- Models: DeepLabv3, DeepLabv3+ with ResNet101 backbone
- Polynomial Learning Rate decay

Experimental Setup

- Datasets: PASCAL VOC 2012, CamVid
- Models: DeepLabv3, DeepLabv3+ with ResNet101 backbone
- Polynomial Learning Rate decay
- Ground-Truth is downsampled to compute loss (memory restrictions)

Datasets

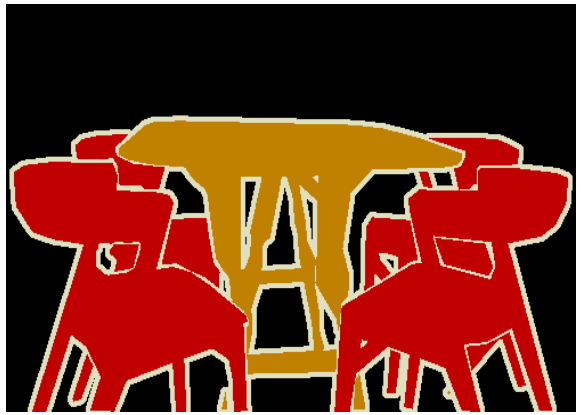
PASCAL VOC 2012

- Web Images
- 20 Classes
- 1464 Training Images

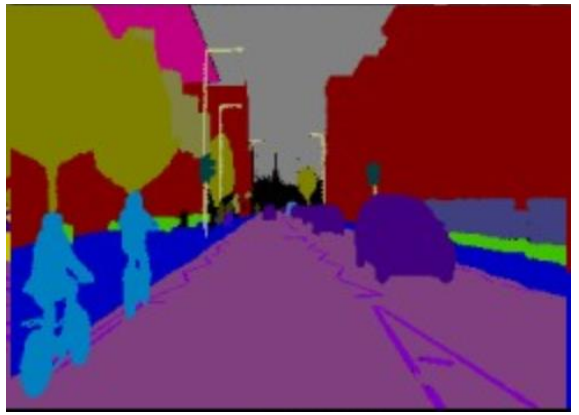
CamVid (Cambridge-driving Video)

- Street Scenes
- 11 classes
- 367 Training Images









Results: PASCAL VOC 2012

(a) **ResNet101-DeepLabv3**

	Method	Inf. Time	mIoU (%)
DeepLabv3	CE* [6]	unknown	77.21
	CE	0.12 s	77.14
	BCE	0.12 s	77.09
	CE & CRF-1	0.39 s	78.32
	CE & CRF-5	0.71 s	78.40
	CE & CRF-10	**1.11** s	78.28
	Affinty	0.12 s	76.24
	RMI	0.12 s	**78.71**

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Results: PASCAL VOC 2012

(a) **ResNet101-DeepLabv3**

Method		Inf. Time	mIoU (%)
DeepLabv3	CE* [6]	unknown	77.21
	CE	0.12 s	77.14
	BCE	0.12 s	77.09
	CE & CRF-1	0.39 s	78.32
	CE & CRF-5	0.71 s	78.40
	CE & CRF-10	**1.11** s	78.28
	Affinty	0.12 s	76.24
	RMI	0.12 s	**78.71**

(b) **ResNet101-DeepLabv3+**

Method		mIoU (%)
DeepLabv3+	CE* [7]	78.85
	CE	78.17
	BCE	77.41
	CE & CRF-1	78.90
	CE & CRF-5	78.75
	CE & CRF-10	78.60
	Affinty	77.09
	RMI	**79.66**

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Results: PASCAL VOC 2012

(a) **ResNet101-DeepLabv3**

	Method	Inf. Time	mIoU (%)
DeepLabv3	CE* [6]	unknown	77.21
	CE	0.12 s	77.14
	BCE	0.12 s	77.09
	CE & CRF-1	0.39 s	78.32
	CE & CRF-5	0.71 s	78.40
	CE & CRF-10	**1.11** s	78.28
	Affinty	0.12 s	76.24
	RMI	0.12 s	**78.71**

+1.5

(b) **ResNet101-DeepLabv3+**

	Method	mIoU (%)
DeepLabv3+	CE* [7]	78.85
	CE	78.17
	BCE	77.41
	CE & CRF-1	78.90
	CE & CRF-5	78.75
	CE & CRF-10	78.60
	Affinty	77.09
	RMI	**79.66**

+0.8

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Results: CamVid

Method		mIOU
DeepLabv3	CE	58.50
	CE + CRF	58.59
	Affinity	59.60
	RMI	**61.11** (+2.61)
DeepLabv3+	CE	62.82
	CE + CRF	61.68
	Affinity	64.42
	RMI	**65.06** (+2.24)

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Ablation Studies: PASCAL VOC 2012

(a) **Downsampling manner and factor**

Method		$\mathcal{DF}$	$\mathcal{R}$	mIoU (%)
DeepLabv3	RMI - Int.	4	3	77.58
	RMI - Max.	4	3	78.62
	RMI - Avg.	4	3	**78.71**
	RMI - Avg.	2	3	**79.23**
	RMI - Avg.	3	3	79.01
	RMI - Avg.	4	3	78.71
	RMI - Avg.	5	3	78.50
	RMI - Avg.	6	3	78.28

(b) **Square region size**

Method		$\mathcal{DF}$	$\mathcal{R}$	mIoU (%)
DeepLabv3	RMI - Avg.	4	1	78.25
	RMI - Avg.	4	2	78.52
	RMI - Avg.	4	3	**78.71**
	RMI - Avg.	4	4	78.65
	RMI - Avg.	6	2	78.06
	RMI - Avg.	6	3	78.28
	RMI - Avg.	6	4	**78.33**
	RMI - Avg.	6	5	78.26

All figures and other images are from the respective papers, copyright rests with original owners or authors.

QUESTIONS ???

All figures and other images are from the respective papers, copyright rests with original owners or authors.

Derivation of RMI Loss Function

Lower bound of Mutual Information

- Normal distribution has maximum entropy among distributions w/ same covariance
- $\mathrm{I(Y;P)} = \mathrm{H(Y)} - \mathrm{H(Y|P)}$

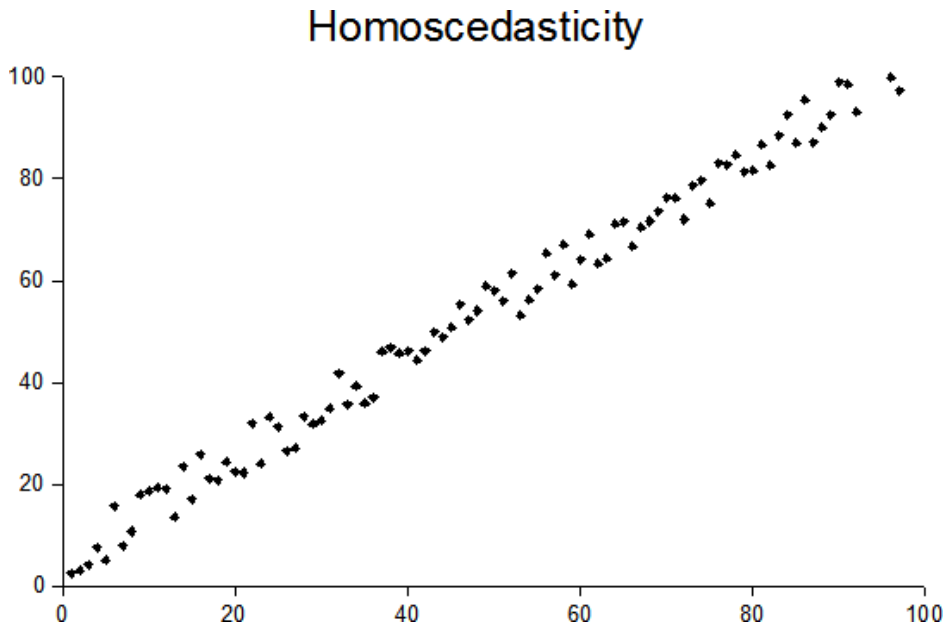
  $>= \mathrm{H(Y)} - \frac{1}{2}\log((2\pi e)^d \det(\Sigma_{Y|P}))$

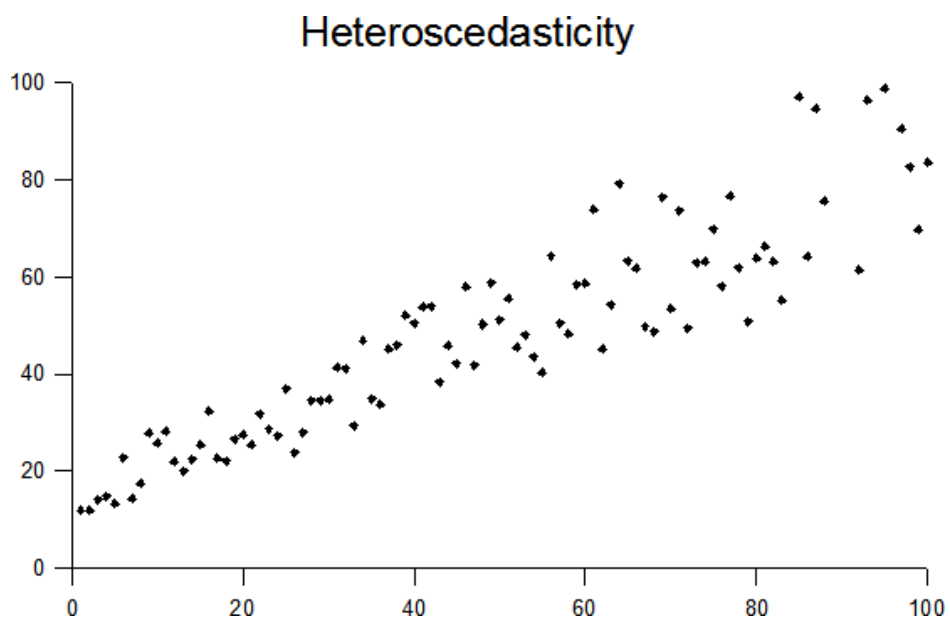
  $>= -\frac{1}{2}\log(\det(\Sigma_{Y|P}))$
- Problem reduced to finding posterior co-variance matrix

Approximation of posterior variance

- Regression matrix of Y on P:
  - $Y = A_{yp}P + \epsilon$
  - $A_{yp} = cov(Y, P)\Sigma_P^{-1}$
- Assumption P and $Y - A_{yp}P$ are second order independent

Intuition for independence assumption





Figures from Wikipedia, copyright rests with original creator(s).

Approximation of posterior variance

- $\mathrm{Var}(Y|P=p) = \mathrm{Var}\left(Y - A_{yp}P\right)$ (Follows from assumption, proof in references)
- $=> \mathrm{Var}(Y|P=p) = \Sigma_Y - \mathrm{Cov}(\mathrm{Y},\mathrm{P})\Sigma_P^{-1^T}\mathrm{Cov}(\mathrm{Y},\mathrm{P})^T$
- Plugging in to original formula for loss:
- $I_L(\mathrm{Y};\mathrm{P}) = -\frac{1}{2}\log(\det(M))$

Further Refinements to RMI loss

- Normalization:
  - $\mathrm{I_L(Y;P)} = -\frac{1}{2\mathrm{d}}\log(\det(\mathrm{M}))$
- Underflow avoidance:
  - $\mathrm{I_L(Y;P)} = -\frac{1}{2\mathrm{d}}\mathrm{Tr}(\log(\mathrm{M}))$