Image Segmentation with Gated Shape CNN for Autonomous Driving

Jeanine Liebold

Intelligent Robotics - 02.12.2019

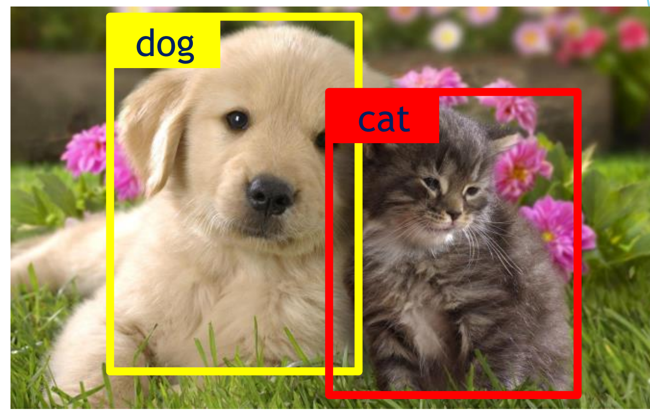
Outline

- Motivation
- Fundamentals
- Gated Shape CNN
- Experiments
- Results
- Conclusion
- References

Motivation

- Image classification
- Object detection
- Image segmentation
  - pixel wise classifiction
  - shape

[6]



[7]

input image

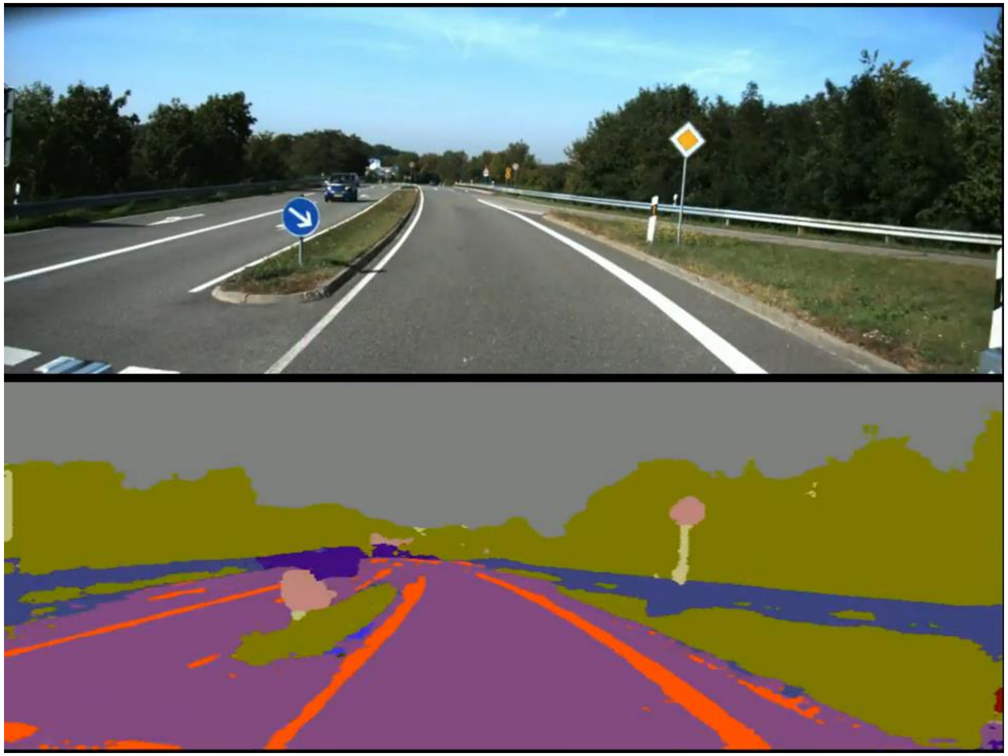
segmentation map

segmentation overlay

[4]

Motivation

Image Segmentation in 2015



[3]

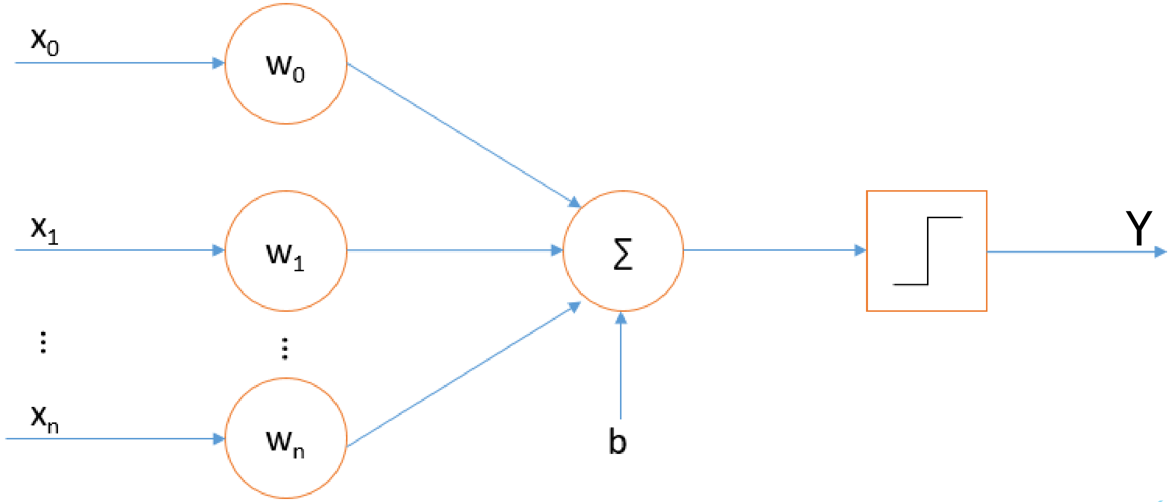
Motivation

Ground-Truth

[2]

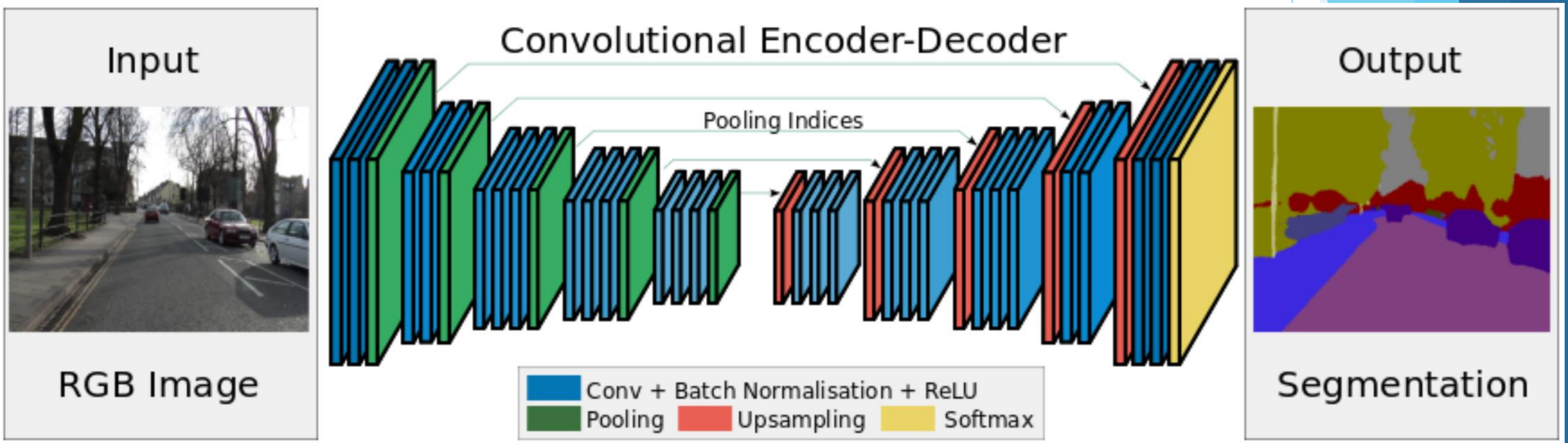
Fundamentals – Neural Networks

- Optimization problem
- All weights initialized randomly
- Loss is calculated (segmentation map/ground-truth)
- Weights optimized based on optimizer



x-input; w-weights; b-bias; y-output

Fundamentals – Convolutional Neural Networks



[3]

Fundamentals – CNN Image Classification

(a) Texture image

81.4%	**Indian elephant**
10.3%	indri
8.2%	black swan

(b) Content image

71.1%	**tabby cat**
17.3%	grey fox
3.3%	Siamese cat

(c) Texture-shape cue conflict

63.9%	**Indian elephant**
26.4%	indri
9.6%	black swan

[5]

- Objects depending more on shape then on texture:
  - small
  - high distance

How to avoid noisy boundaries and loss of detail in high distances?

Gated Shape CNN

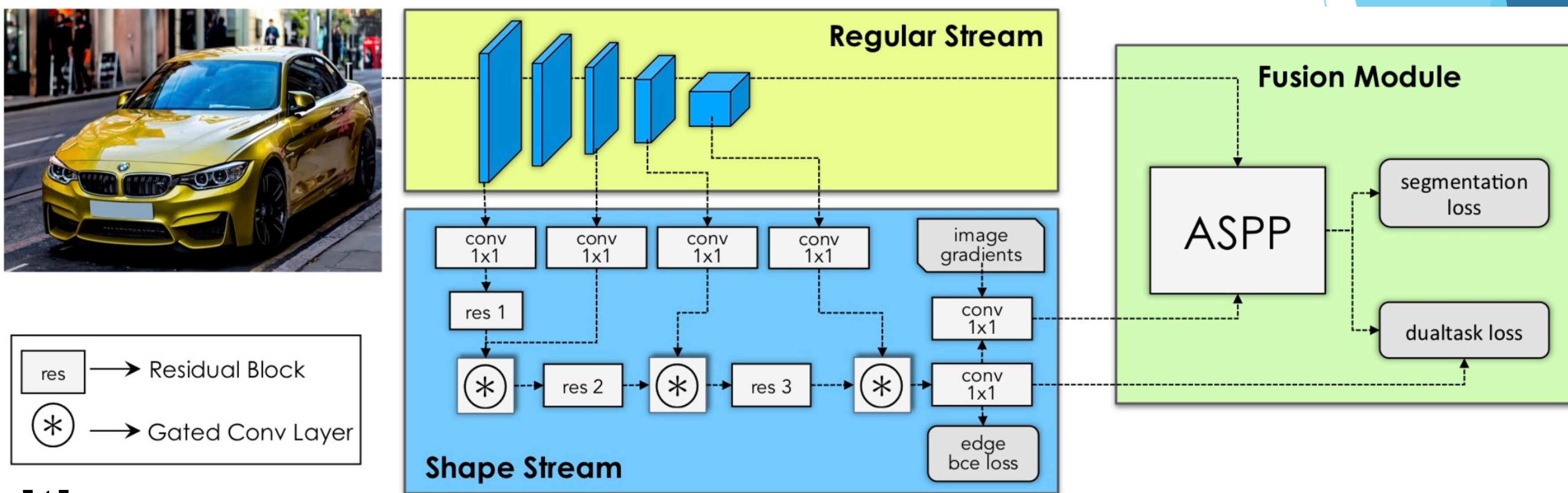
- Title of Paper: "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation"
- Authors:
  - Towaki Takikawa (NVIDIA)
  - David Acuna (University of Waterloo)
  - Varun Jampani (University of Toronto)
  - Sanja Fidler (Vector Institute)
- Published: 12 July 2019, ICCV 2019

Gated Shape CNN - Approach

- Seperate color, texture and shape processing
- Information gets fused in very top layer
- New type of gates in architecture
- Cityscape dataset:

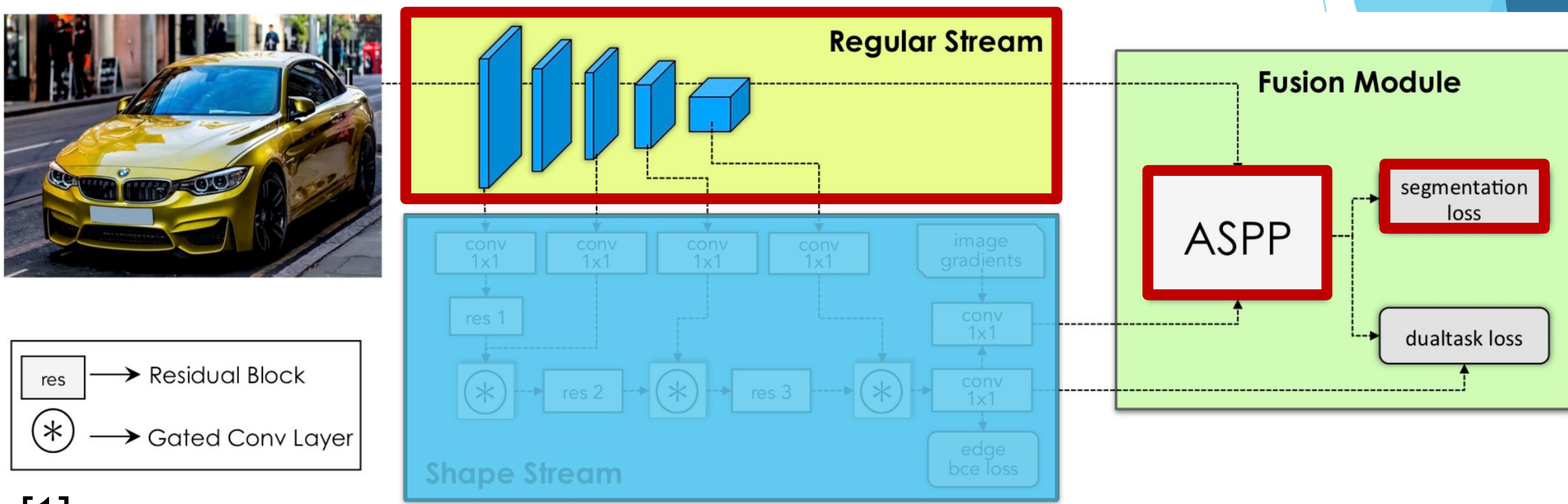
[3]

Gated Shape CNN – Architecture



[1]

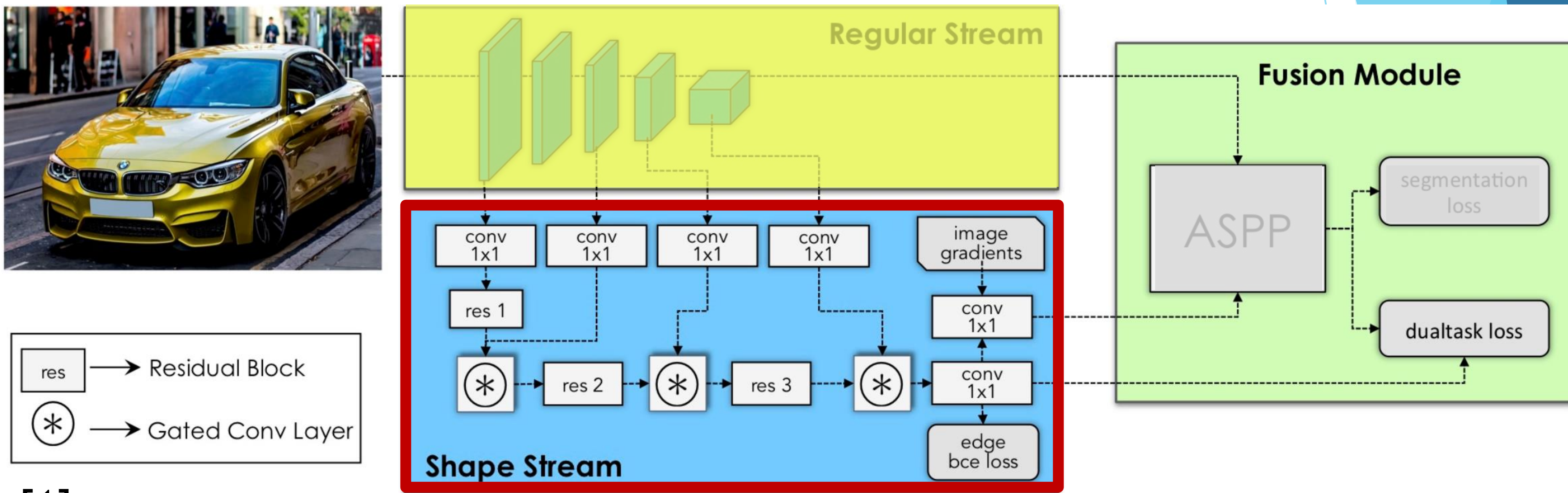
Gated Shape CNN – Architecture



[1]

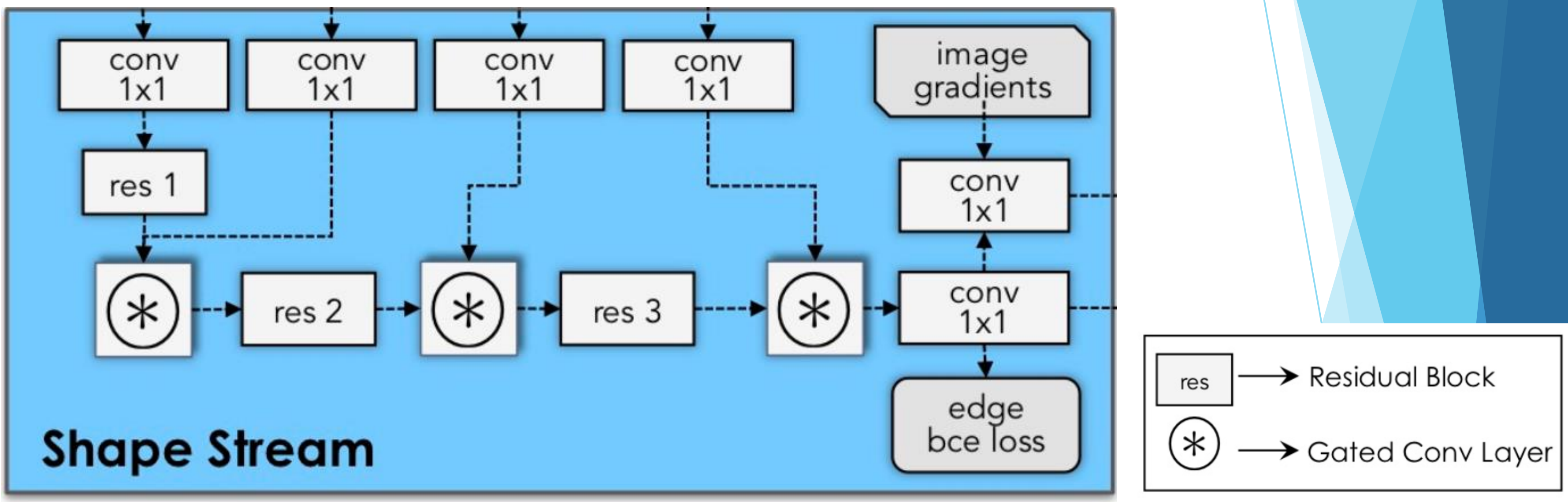
e.g. DeepLabV3+ (Google)

Gated Shape CNN – Architecture



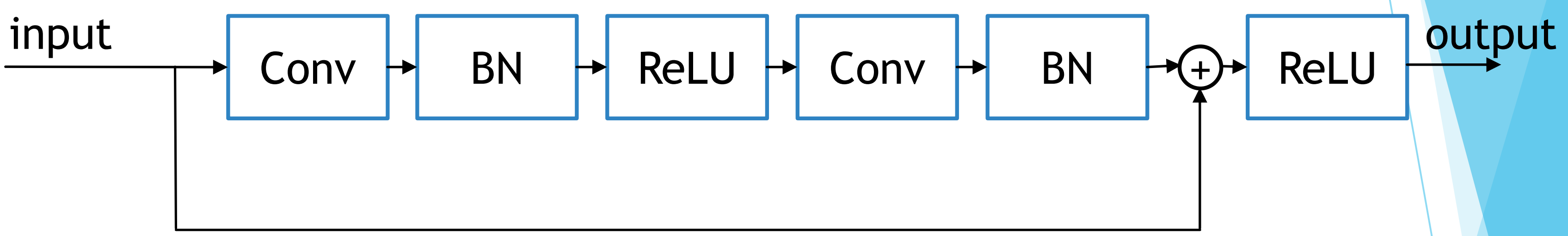
[1]

Gated Shape CNN – Shape Stream



[1]

Gated Shape CNN – Shape Stream (Residual Block)

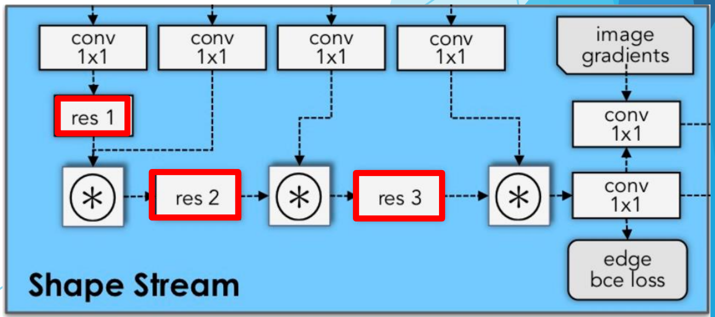


Conv: Convolution

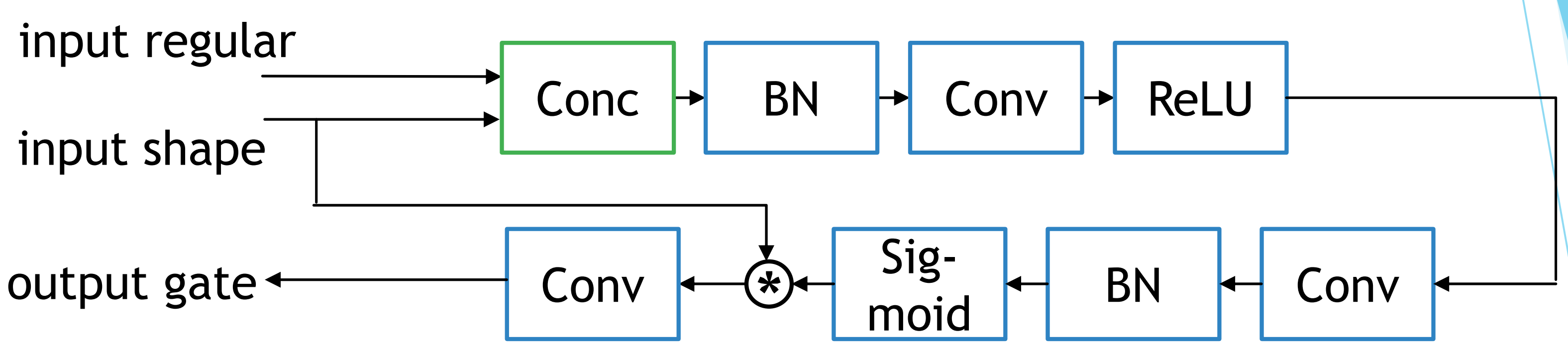
BN: Batch Normalization

ReLu: Activation with Rectifier Linear Unit

[1]



Gated Shape CNN – Shape Stream (Gate)



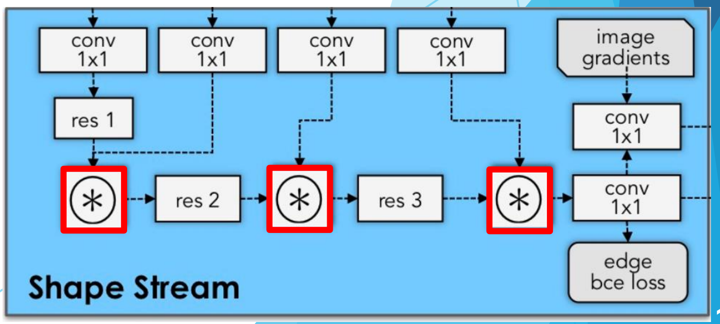
Conv: Convolution

BN: Batch Normalization

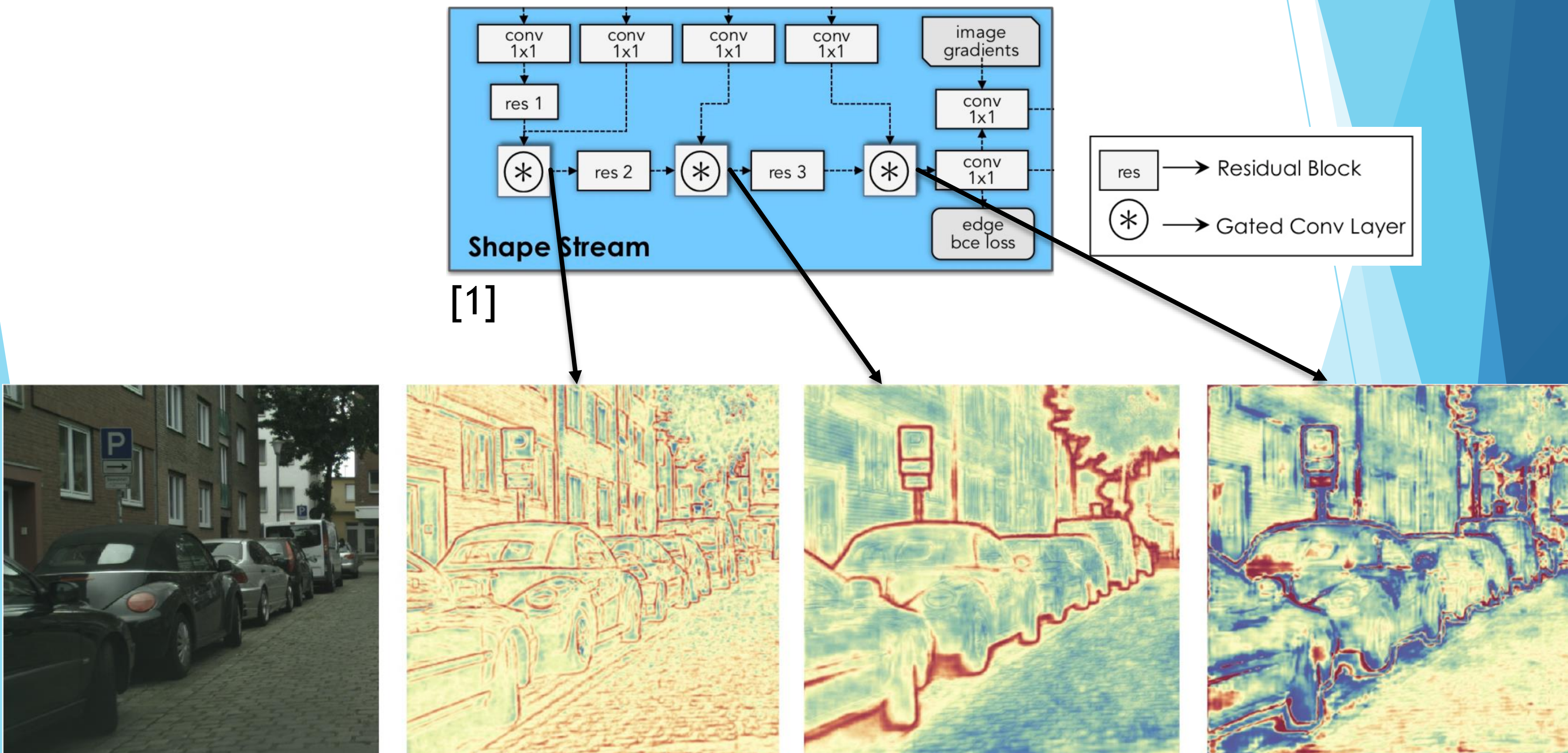
ReLu: Activation with Rectifier Linear Unit

Conc: Concatenation

[1]



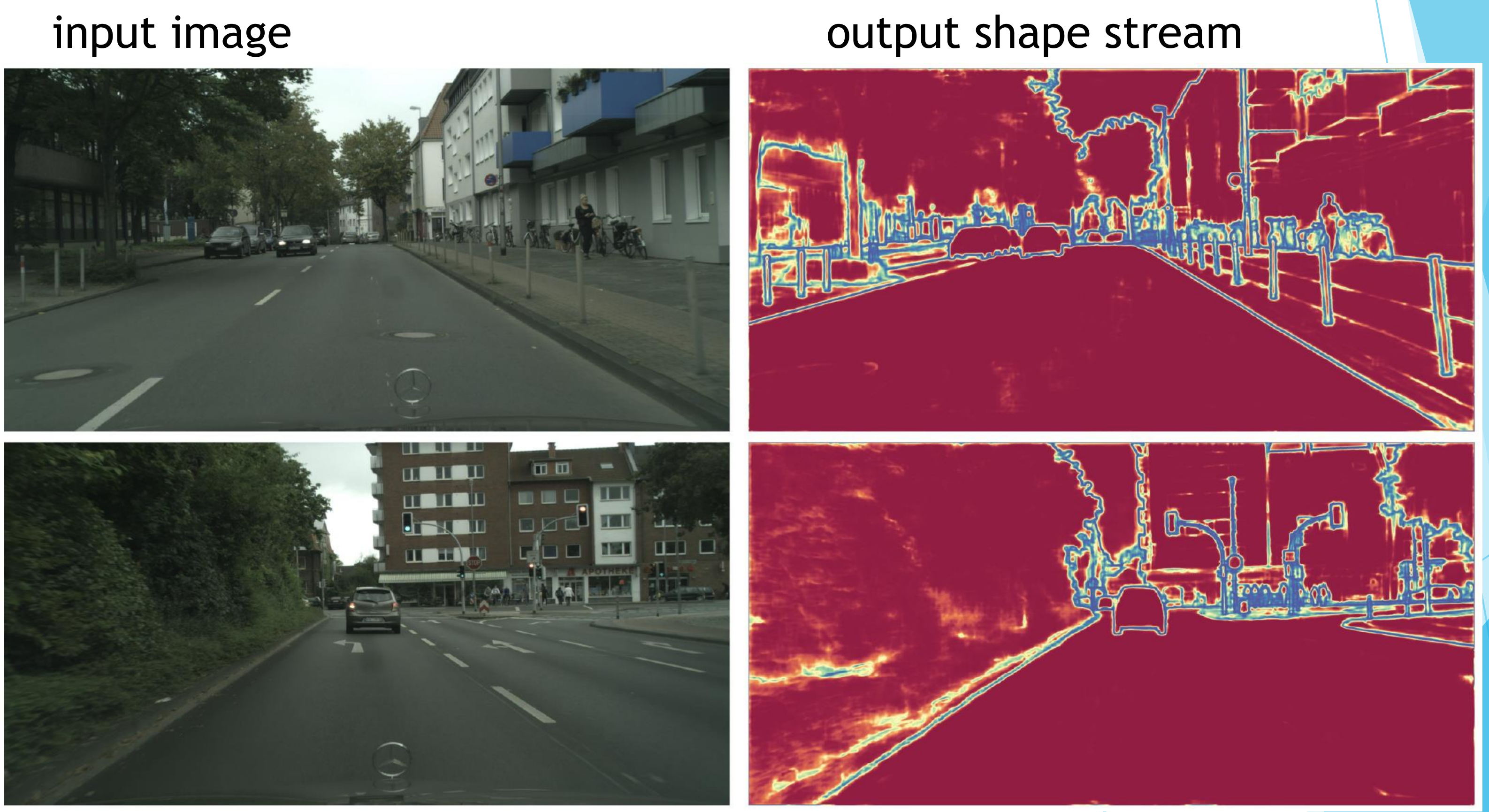
Gated Shape CNN - Output Gates 1-3



[1]

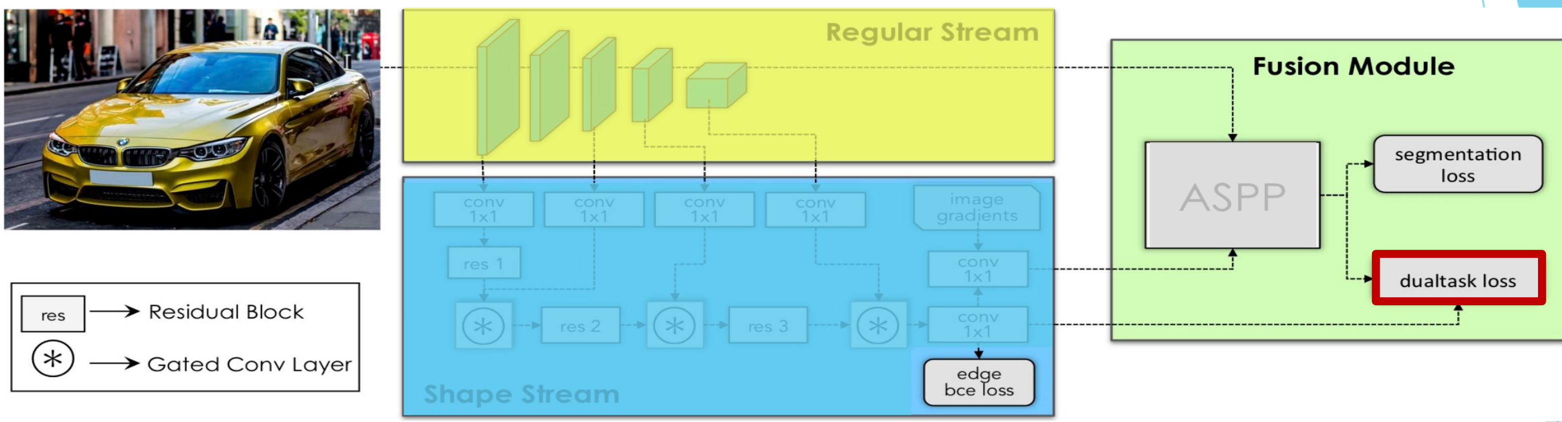
[1]

Gated Shape CNN - Output Shape Stream



[1]

Gated Shape CNN – Dual Task Loss

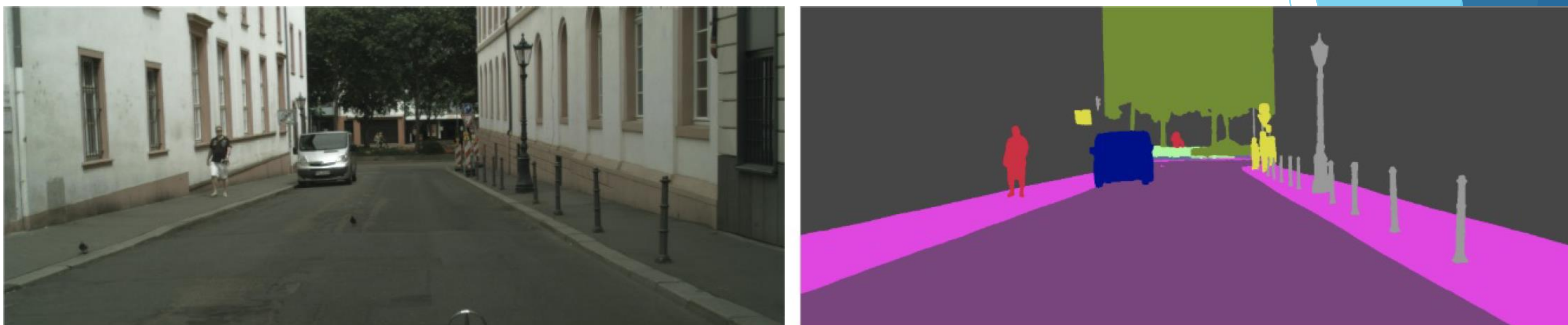


[1]

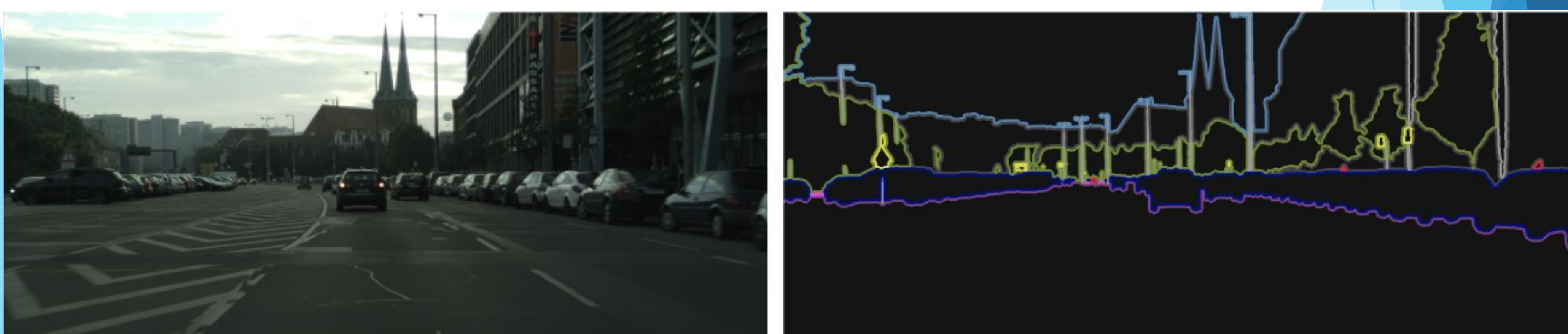
- Combination of the two loss functions
  - semantic segmentation
  - boundary segmentation

Experiments

- Segmentation mask



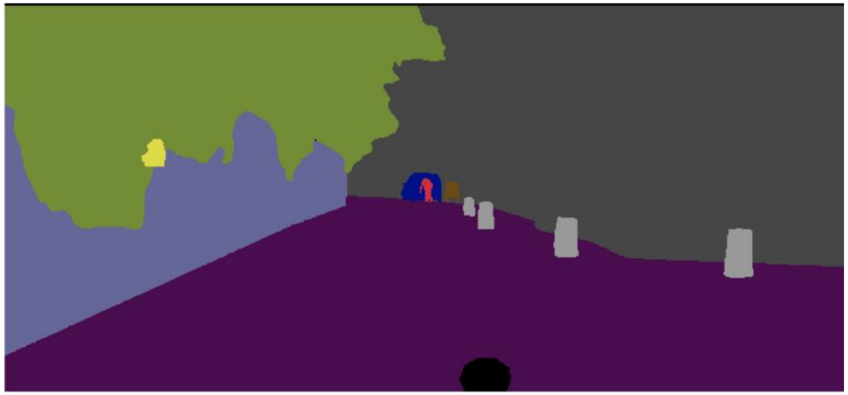
- Boundaries of predicted segmentation masks



[1]

Experiments

- Distance based evaluation
- Mulitple crop factors



Crop Factor: 0

Crop Factor: 400

[1]

Results –
Errors in Predictions

[1] original ground-truth DeepLabV3+ Gated SCNN

Results – Evaluation

- Baseline – DeepLabV3+
- Evaluation Metrics
  - IoU = $\frac{TP}{TP+FP+FN}$ = intersection over union
  - F-score along the boundary
    - $\frac{TP}{TP+FP} \triangleq$ precision
    - $\frac{TP}{TP+FN} \triangleq$ recall
  - F-Score = $\frac{2*recall*precision}{recall+precision}$

TP = true positive pixels
FP = false positive pixels
FN = false negative pixels

Results – Intersection over Union (IoU)

Method	pole	t-light	t-sign	person	rider	mean
LRR	58.6	68.2	72.0	81.6	60.0	69.7
DeepLabV2	49.6	57.9	67.3	79.8	59.8	70.4
Piecewise	51.1	65.0	71.7	81.5	61.1	71.6
PSP-Net	62.6	71.8	80.7	82.1	61.5	78.8
DeepLabV3+	65.2	68.6	78.9	82.3	62.8	78.8
Ours	**70.8**	**75.9**	**83.1**	**85.3**	**67.9**	**80.8**

[1]

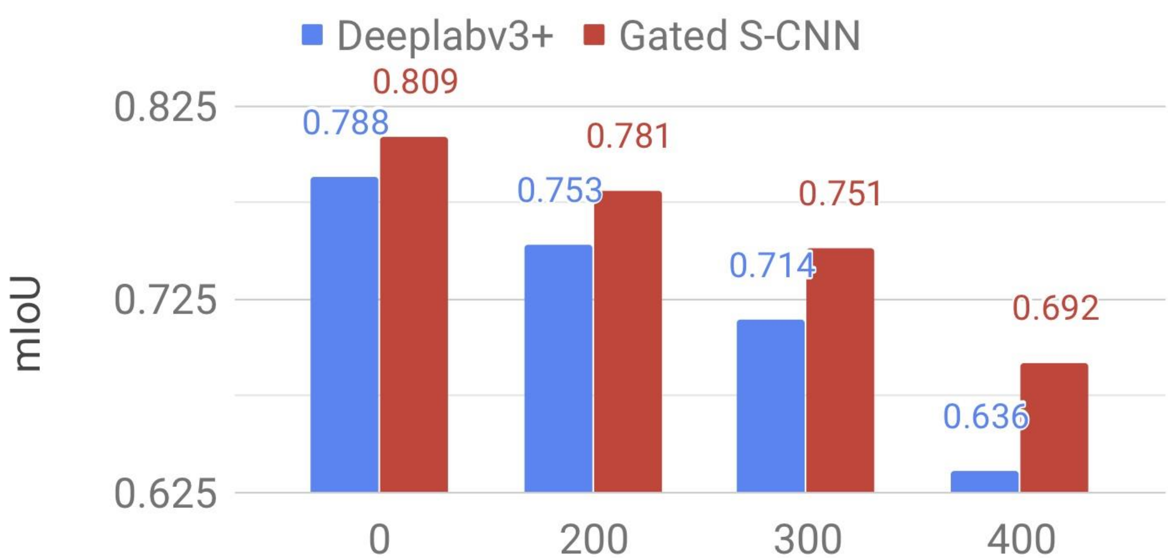
Results –
Boundary F-Score

Thrs	Method	pole	t-light	t-sign	person	rider	mean
12px	DeepLabV3+	83.8	75.2	81.2	76.6	78.7	80.1
	Ours	**87.1**	**82.3**	**84.4**	**80.4**	**82.8**	**81.8**
9px	DeepLabV3+	82.1	73.7	79.5	74.7	76.8	78.7
	Ours	**86.1**	**81.5**	**83.3**	**79.1**	**81.5**	**80.7**
5px	DeepLabV3+	77.9	69.0	74.7	69.0	71.9	74.7
	Ours	**83.6**	**78.6**	**80.4**	**75.4**	**77.8**	**77.6**
3px	DeepLabV3+	72.0	62.8	67.7	61.5	66.4	69.7
	Ours	**79.6**	**74.3**	**76.2**	**69.8**	**73.1**	**73.6**

[1]

Results –
Different Crop Factors

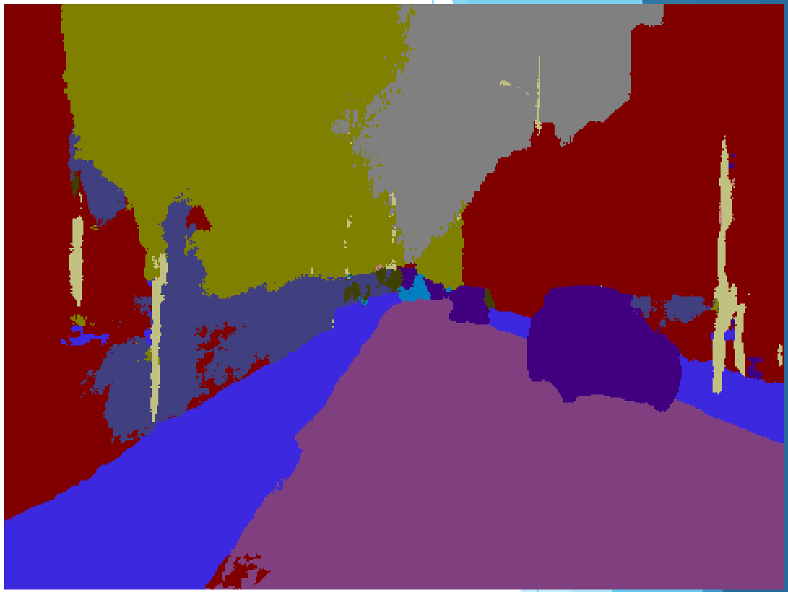
- Mean intersection over union (mIoU)



[1]

Conclusion

[1] GSCNN (2019)

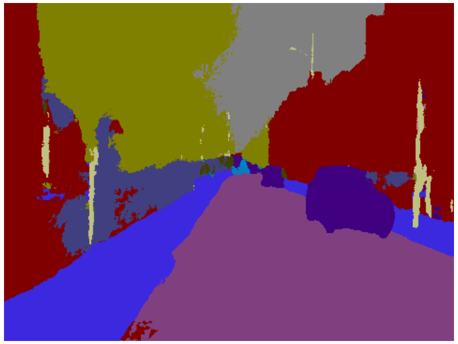


[3] SegNet (2015)

How to avoid
noisy boundaries and
loss of detail in high distances?

Conclusion

[1] GSCNN (2019)



[3] SegNet (2015)

- Two-Stream CNN architecture leads to:
  - sharper predictions around object boundaries
  - a boosts performance on thinner and smaller objects
  - crop mechanisms showed improvement in high distance objects

References

[1] Towaki Takikawa, David Acuna, Varun Jampani, and SanjaFidler; **Gated-SCNN: Gated Shape CNNs for semantic segmentation**; ICCV 2019; https://arxiv.org/pdf/1907.05740.pdf, retrieved 29.11.2019

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, Bernt Schiele; **The Cityscapes Dataset for Semantic Urban Scene Understanding**; CVPR 2016, https://www.cityscapes-dataset.com/ retrieved 29.11.2019

[3] Vijay Badrinarayanan, Ankur Handa, Roberto Cipolla; SegNet: **A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling**; CVPR 2015; http://mi.eng.cam.ac.uk/projects/segnet/ retrieved 29.11.2019

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, andHartwig Adam; **Encoder-Decoder with Atrous SeparableConvolution for Semantic Image Segmentation**; ECCV 2018; https://arxiv.org/pdf/1802.02611.pdf retrieved 29.11.2019

References

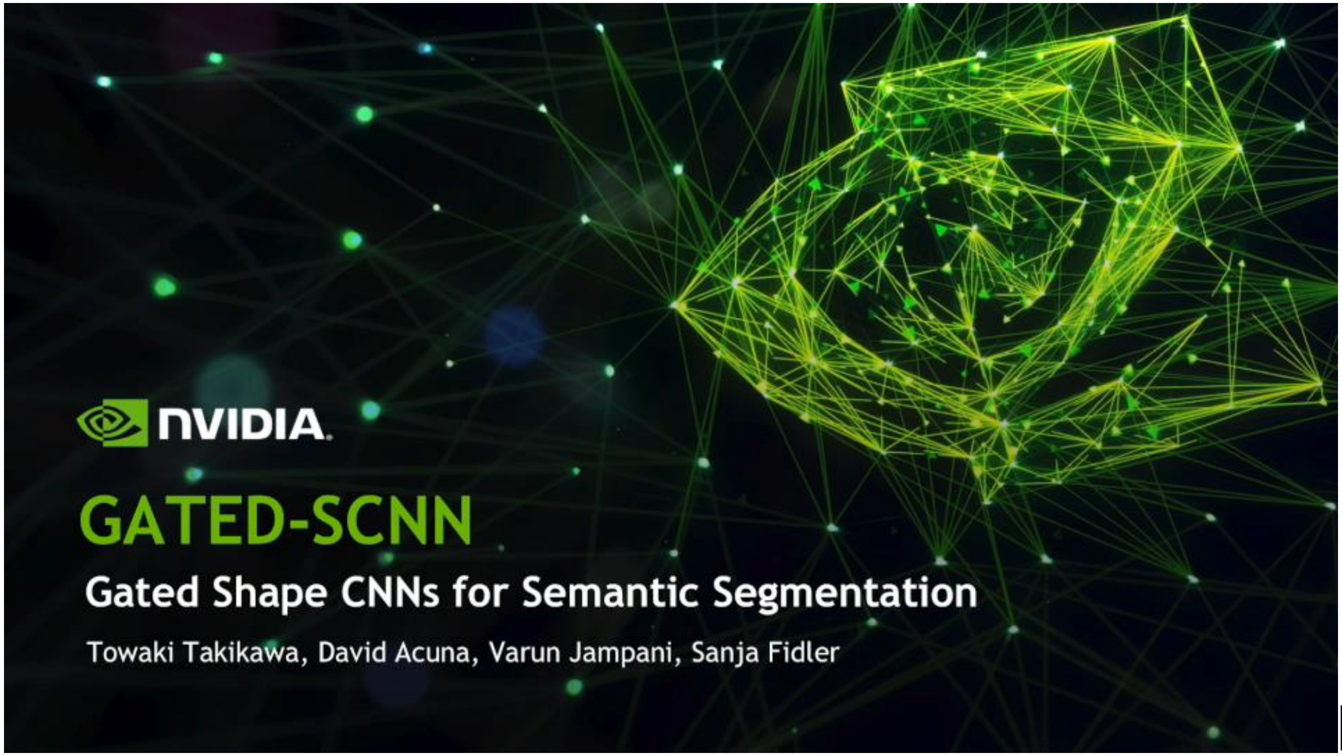
[5] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, Wieland Brendel; **ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness**; ICLR 2019; https://arxiv.org/pdf/1811.12231.pdf retrieved 28.11.2019

[6] Cat image: https://www.cats.org.uk/media/2197/financial-assistance.jpg?width=1600, retrieved 20.11.2019

[7] Dog/cat image: https://i.pinimg.com/originals/1d/c9/ca/1dc9caf8c7ede4c33156bbcaa5edbaba.jpg retrieved 20.11.2019

Github Gated Shape CNN: https://github.com/nv-tlabs/gscnn

Results


NVIDIA
GATED-SCNN
Gated Shape CNNs for Semantic Segmentation
Towaki Takikawa, David Acuna, Varun Jampani, Sanja Fidler

[1]