# OCONet: Image Extrapolation by Object Completion

*Jihyun Kim*
*Vision & Display Systems Lab.*
*Dept. of Electronic Engineering, Sogang University*

# Outline

- Introduction
  - Outpainting task
  - Outpainting trend
- Background
  - GAN
  - PatchGAN Discriminator
  - Encoder Decoder
- OCONet: Image Extrapolation by Object Completion (CVPR 2021)

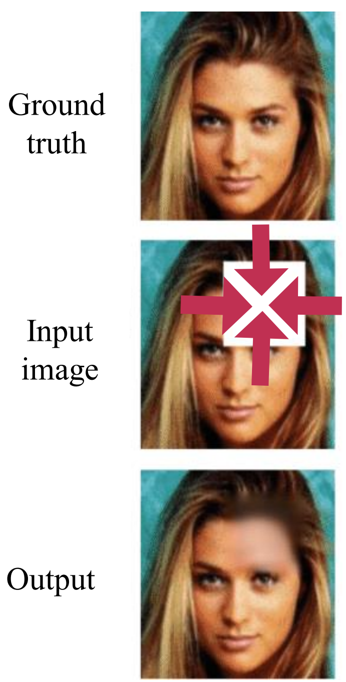
# **Introduction** - outpainting task

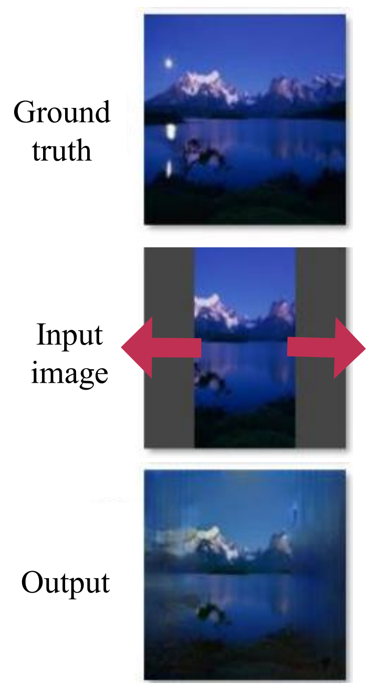
- Generation of outer area of input image
- Major goal - generation of perceptually natural image

# **Introduction** - outpainting task

- Inpainting
  - generation/restoration of inner masked area of an input image
  - Relatively less difficult

- Inpainting
  - Sources from multiple directions



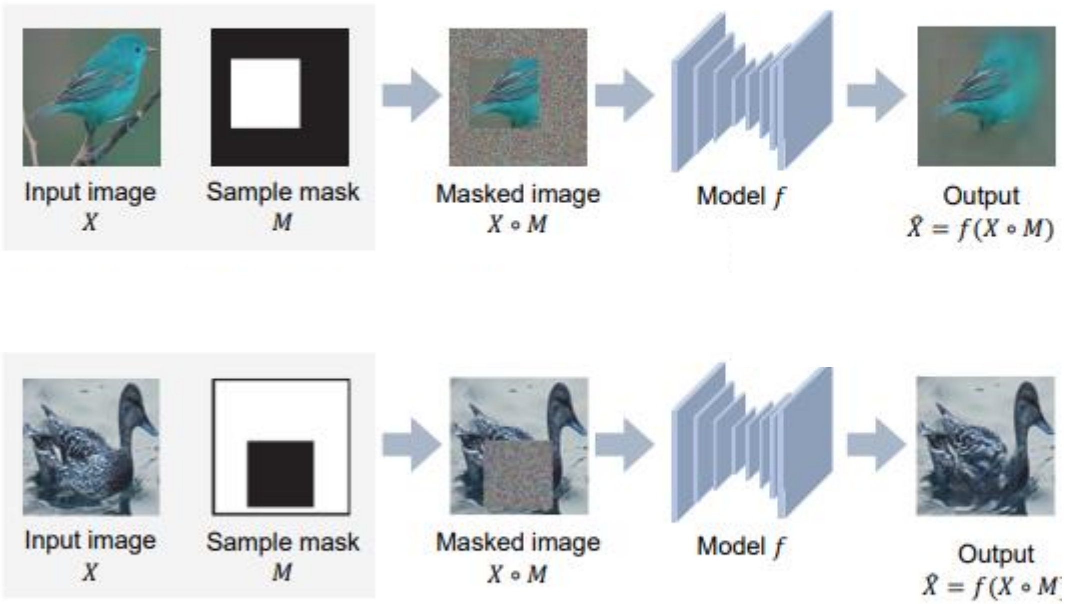

- Outpainting
  - Sources from a single direction

# **Introduction** - outpainting task

- 한계: Object 복원, bluriness

Outpainting

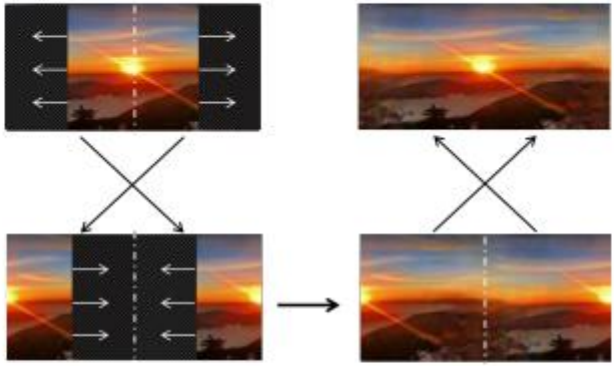
Inpainting



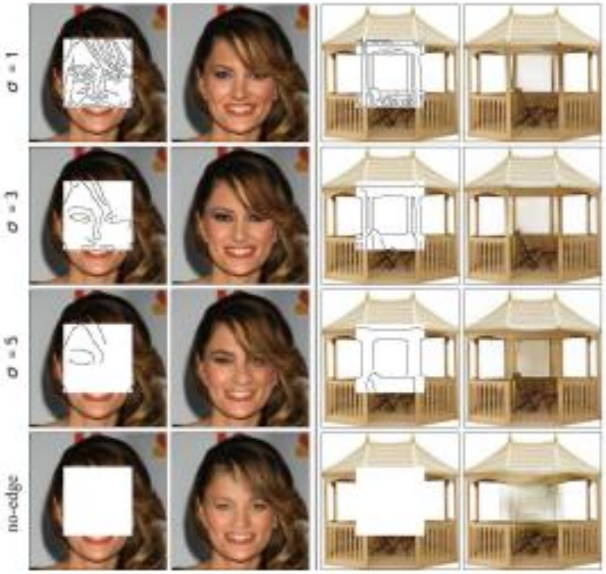

1142.pdf (bmvc2021-virtualconference.com)

# **Introduction -** Outpainting/Inpainting trend

- GAN based method
- provide additional information



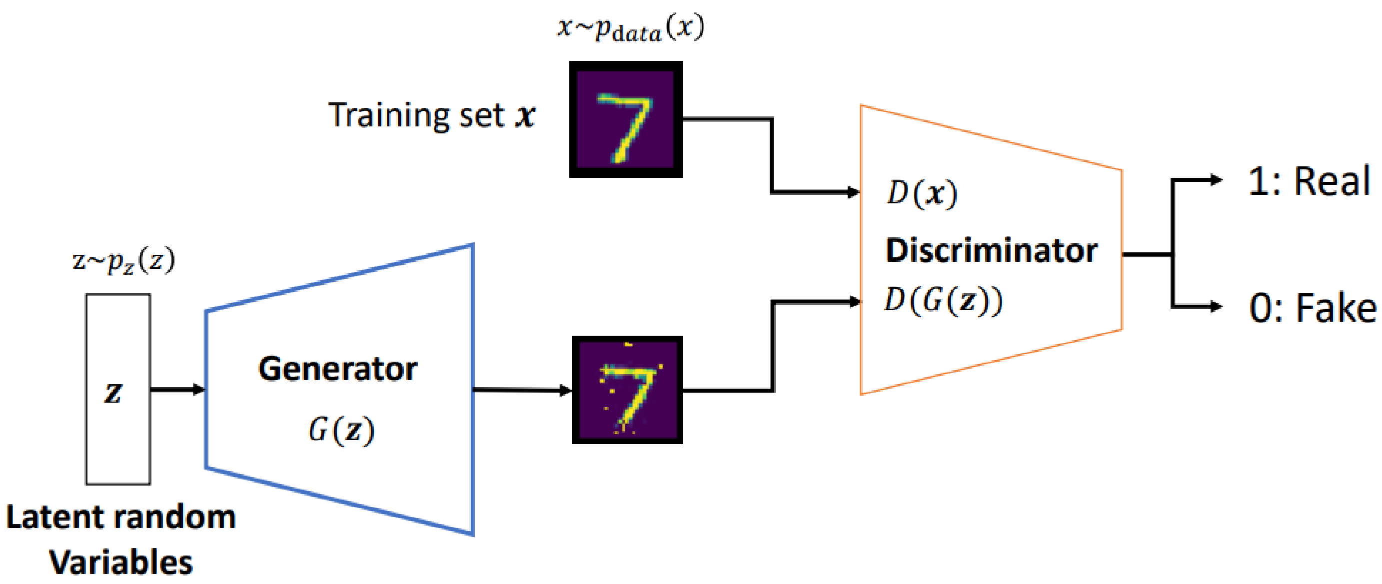
Guided Image Outpainting via BidirectionalRearrangement with Progressive Step Learning (IEEE 2021)



EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning (arXiv 2019)

# **Background -** GAN (Generative Adversarial Networks)

- image generation model
- adversarial learning between Generator and Discriminator

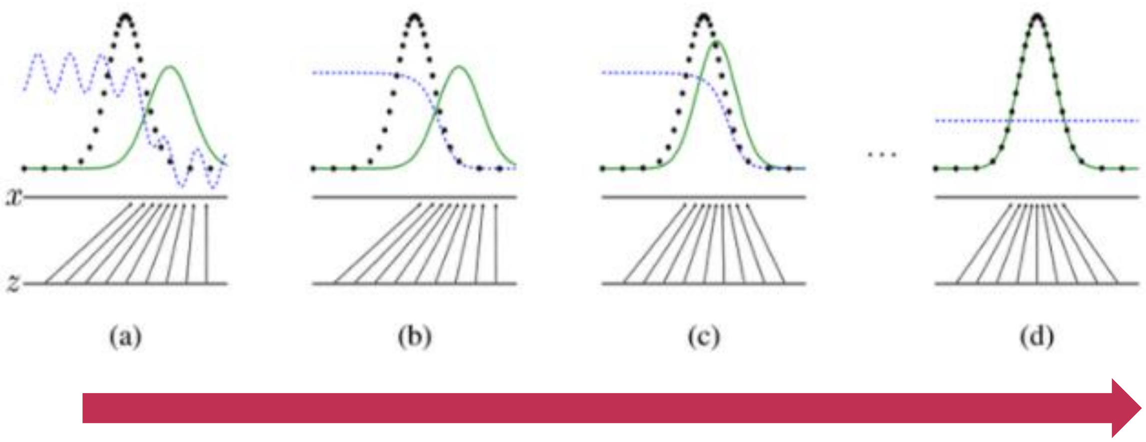


# **Background -** GAN (Generative Adversarial Networks)

- Objective function

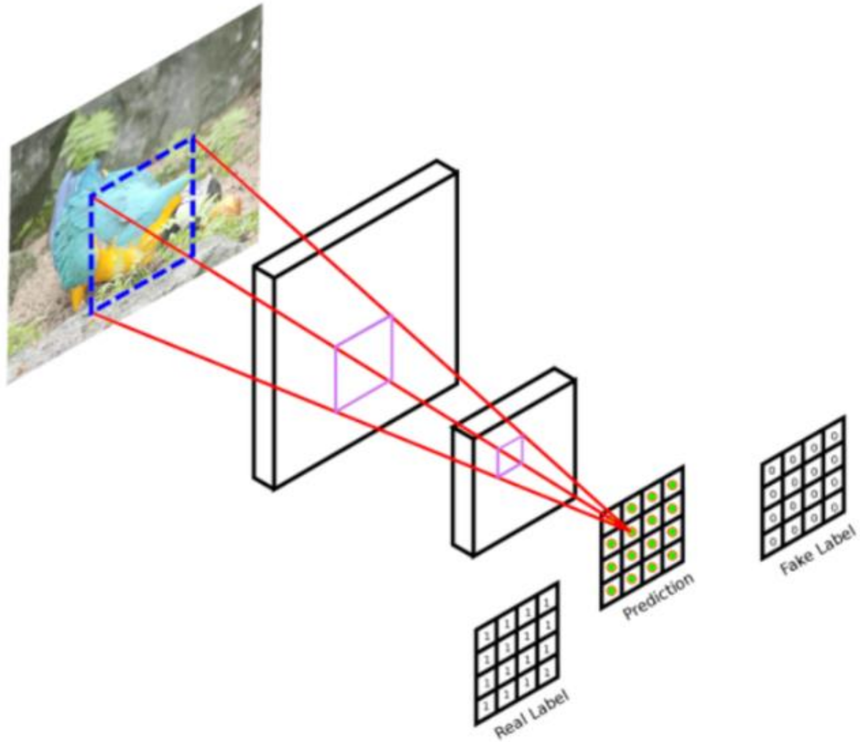
$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

- Training Obejective
  - Generator의 분포가 training data의 분포를 근사



# **Background -** PatchGAN Discriminator

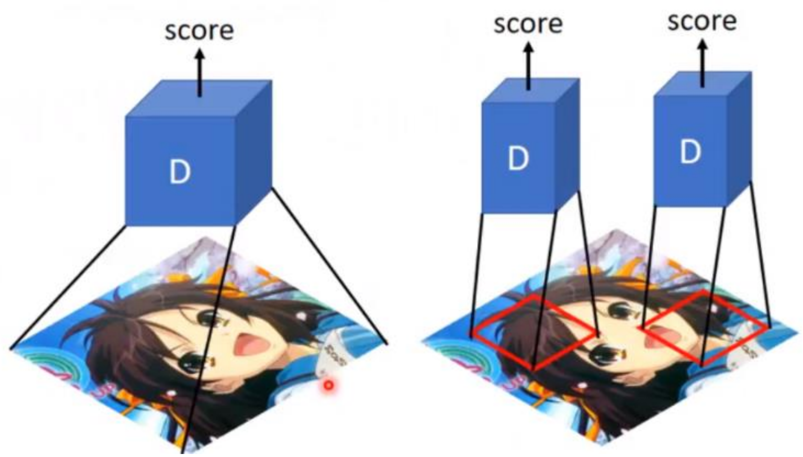
- Whereas origianal discriminator does classification on a whole input image, PatchGAN discriminator does classification on patches of an input image




output: a matrix of values, each between one(Real) and zero(Fake)

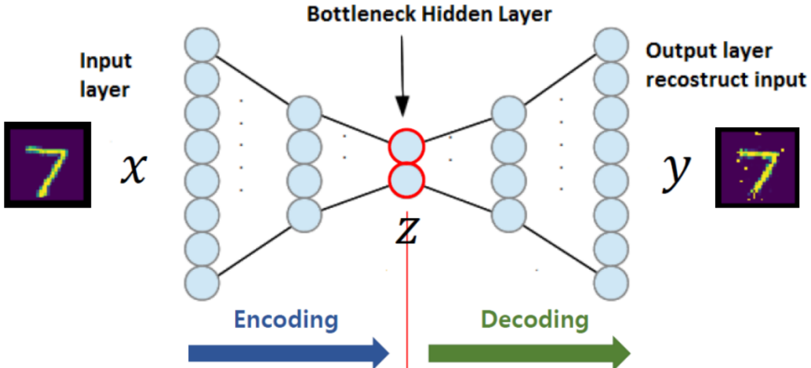
# **Background** - PatchGAN Discriminator

- L1, L2 loss의 한계점을 보완 → low frequency feature 학습 → blurry한 이미지 생성
- Whole image에서 general feature(low-frequency)가 아닌 n x n 으로 쪼갠 patch에 대하여 True/False를 예측
- 이를 통해 Generator가 High-frequency feature(detail)에 대하여 학습 가능



# **Background** - Encoder Decoder

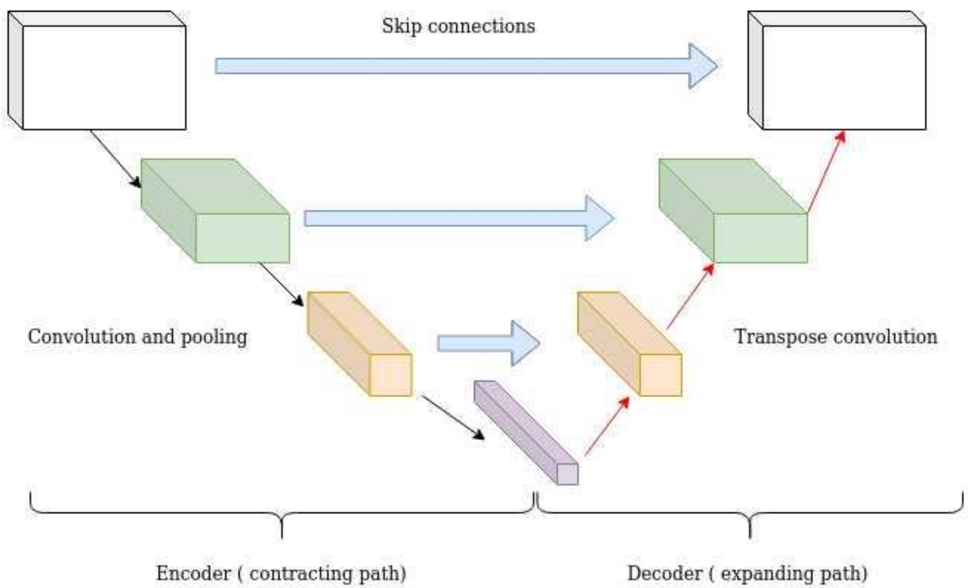
- composed of an Encoder(compresses input data) and a Decoder(reconstructs image)
- used for compressing data and extracting important features
- Latent Variable is usually in lower dimension than the input data




- Code
- Latent Variable (잠재변수)
- Feature
- Hidden representation

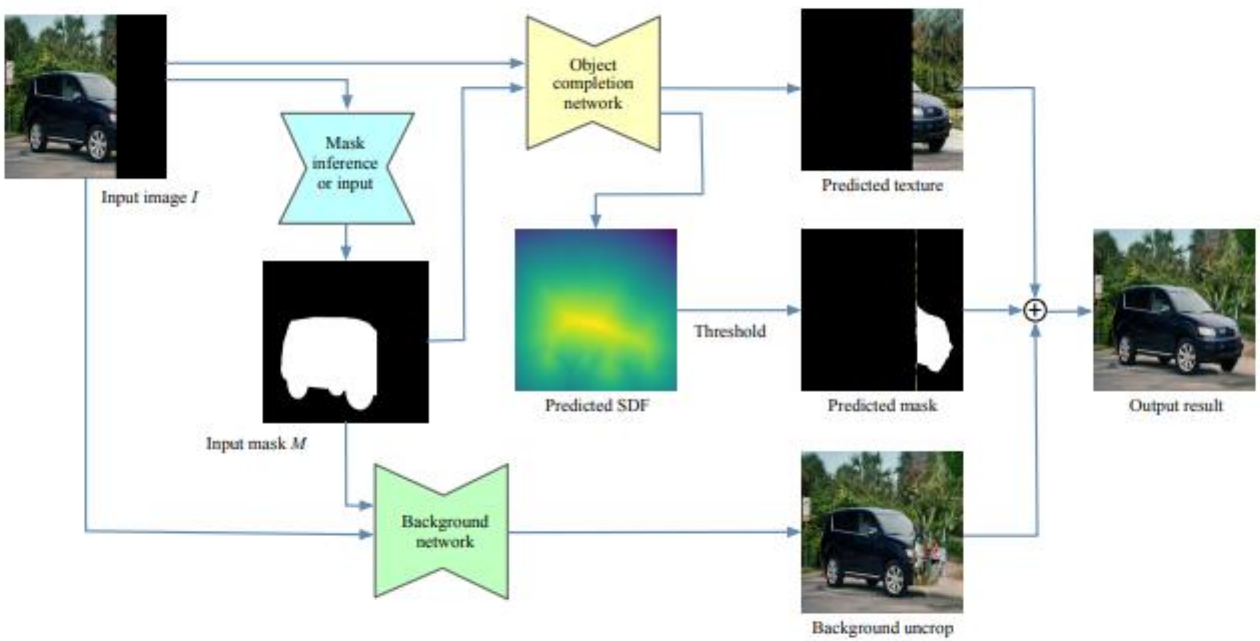
# **Background -** Encoder Decoder with Skip Connection

- 단순 encoder-decoder architecture에서 image가 convolution을 거치며 low level feature가 누락
- Skip connection은 이러한 feature가 누락되기 전에 decoder 또는 뒤쪽의 layer에 전달될 수 있도록 함



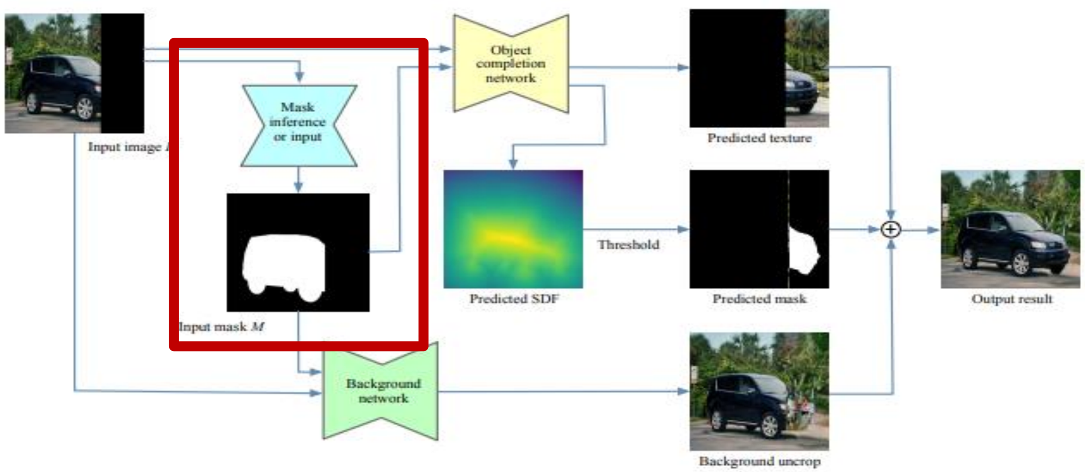
# OCONet - overview

- Major Contributions
  - Object completion network independent from the rest of the extrapolation problem
  - Signed Distance Field as mask representation



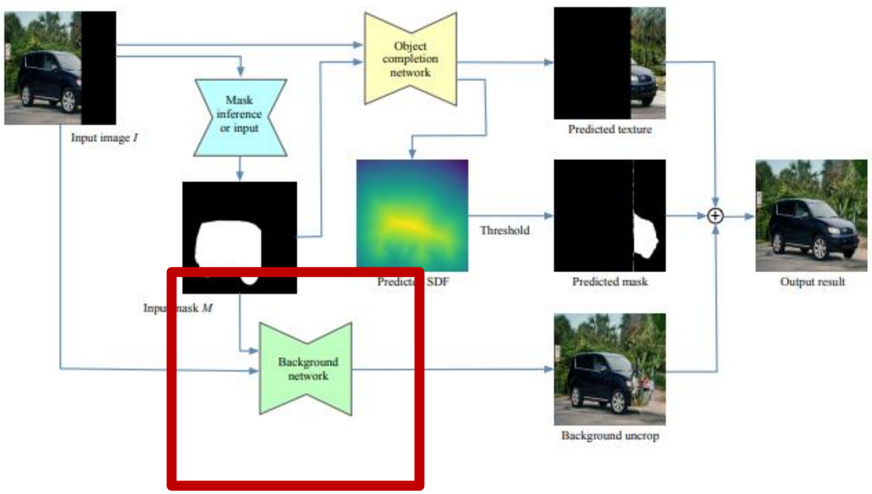
# OCONet – step1 mask generation

- Mask Generation
  - Input: cropped color image I of size H*W*3
  - Output: cropped mask of size H*W*1
- Any instance segmentation method possible (in experiment shape mask)
- External masks can be used



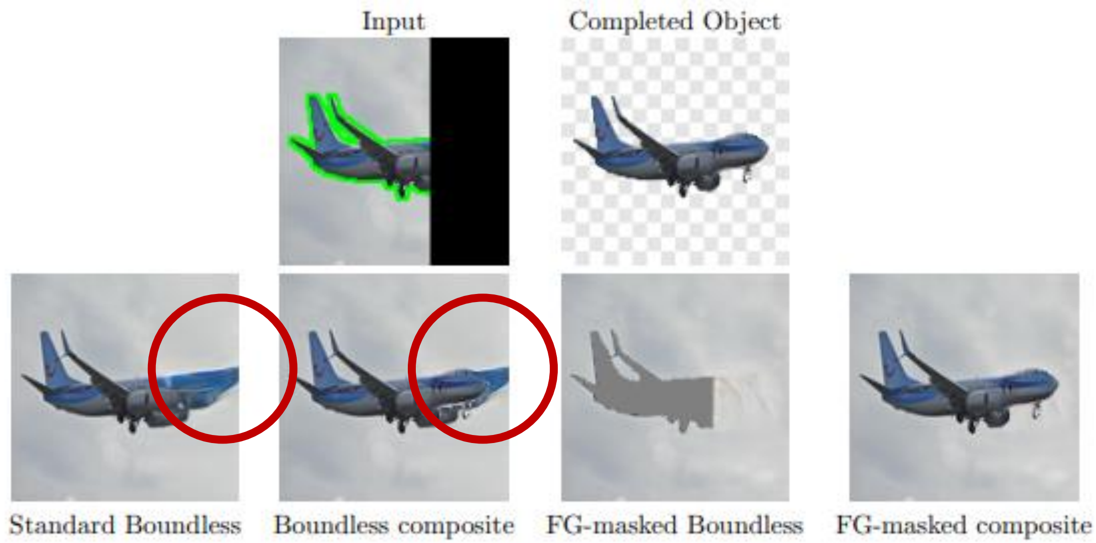
# OCONet – step2 Background extrapolation

- Input:
  - 1) cropped color image I of size H*W*3
  - 2) Input Mask
  - zero out the foreground object in the generator's input using the object mask M.
- Output: background extrapolated image
- Any extrapolation method possible (in experiment Boundless GAN)



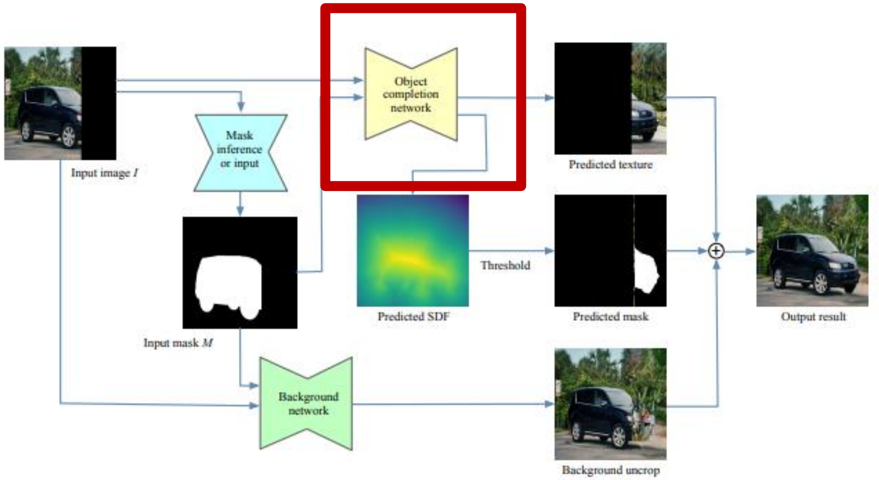
# **OCONet** – step2 Background extrapolation

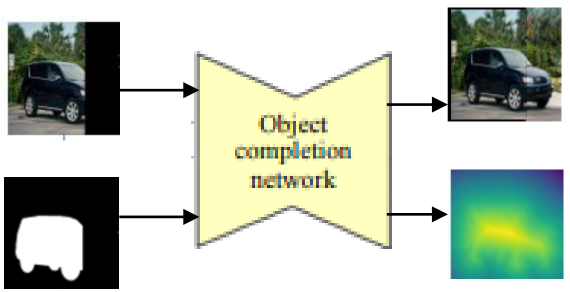
- Masking input image I
  - Pixels sometimes going beyond the extent of Completed Object
  - Better compositing result



# OCONet – step3 main

- Input: [I,M]
  - Input image I
  - Signed Distance Field (Input mask M)
- Output: [predicted texture, predicted mask]
  - Completed object pixels
  - Completed mask represented as Signed Distance Field

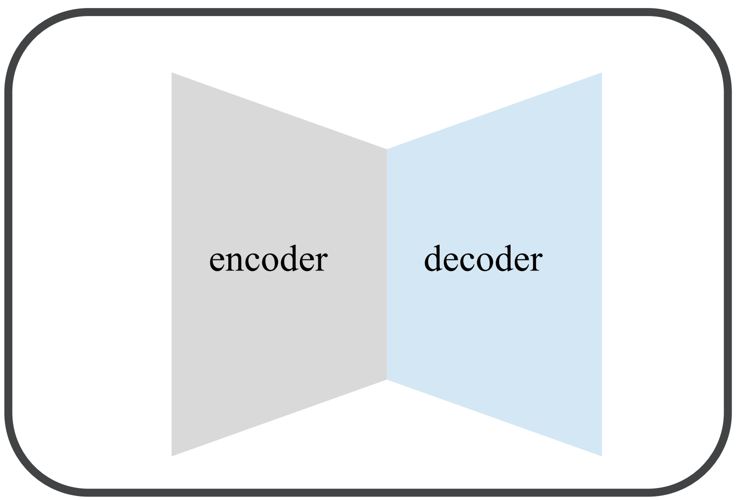


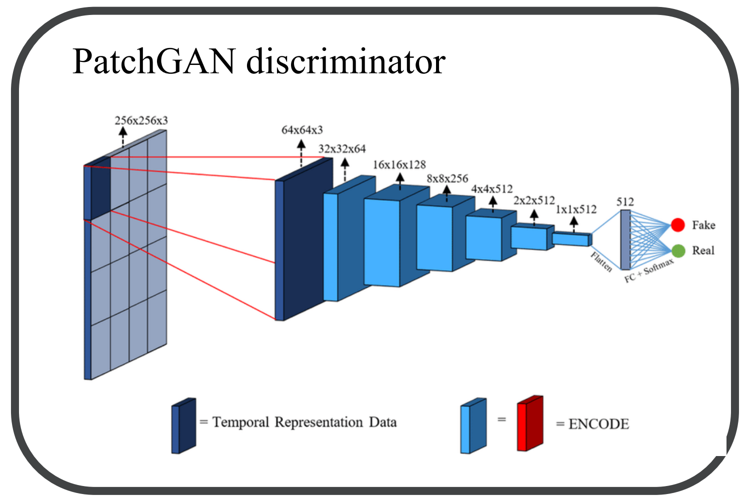
# OCONet – Framework

- Generator: Encoder-Decoder with skip connections
- Discriminator: PatchGAN discriminator

**Generator**

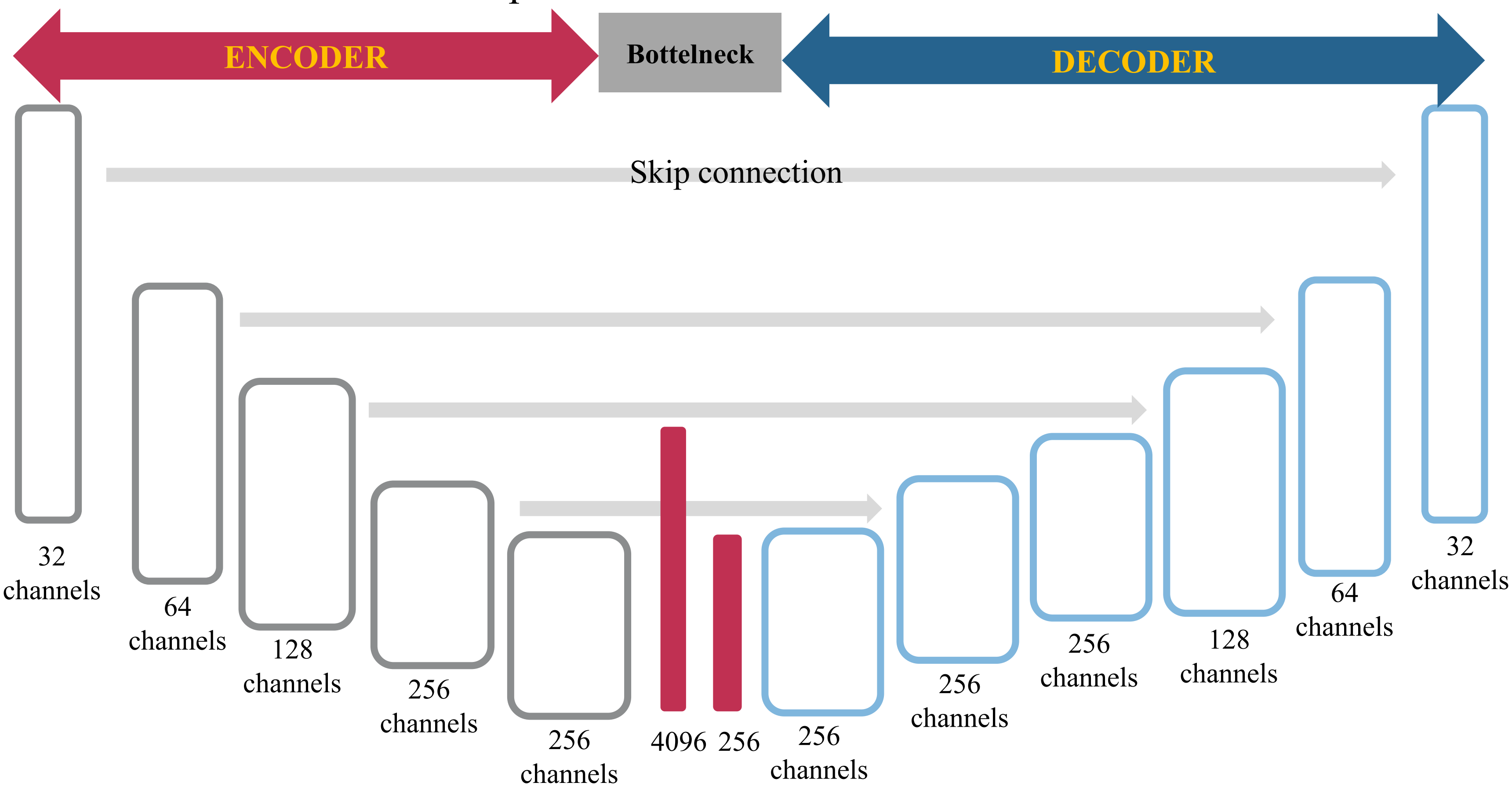



**Discriminator**



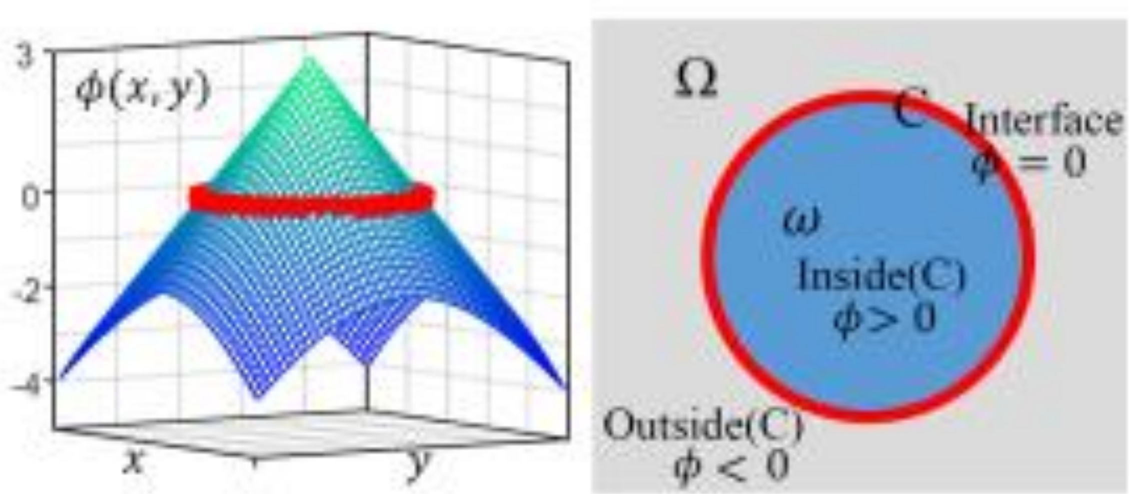
# OCONet – Framework

- Encoder-Decoder with skip connections



# OCONet – step3 main

- Signed Distance Field (SDF)
  - Pixel의 위치를 boundary로부터 pixel거리로 측정하여 pixel 값으로 설정
  - Boundary에 해당하는 pixel들을 0, boundary 내의 pixel들은 +, 밖의 pixel들을 – 값을 가짐



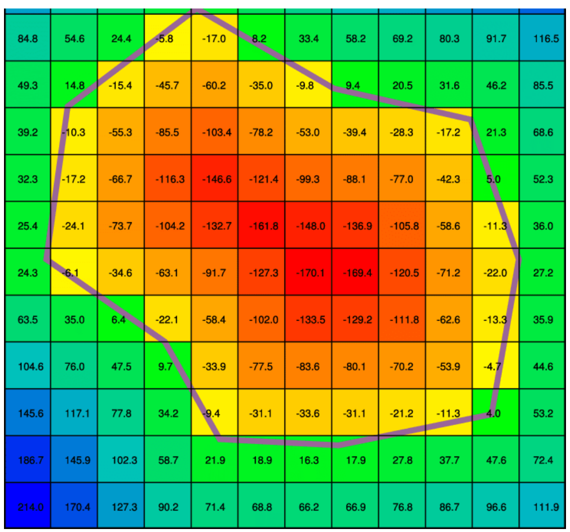



서강대학교 SOGANG UNIVERSITY

VDS LAB

# OCONet – step3 main

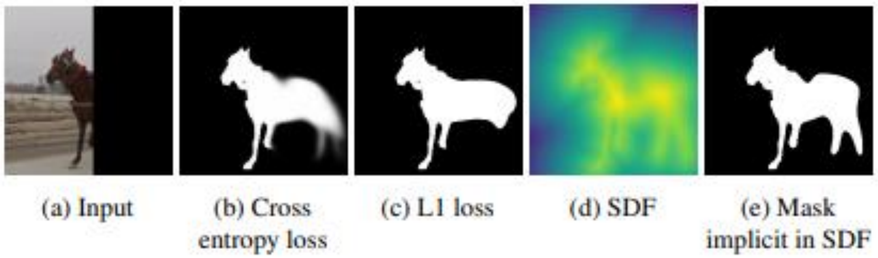
- Signed Distance Field (SDF)
  - Signs indicate the segmentation (0 being the boundary and positive pixels indicating the object) → network is trained to learn the level sets for salient objects
  - Absolute values change gradually → network learns the gradual change around boundaries more naturally
  - Achieves sharpness by thresholding




$$f(x) = \begin{cases} \min_{s \in S} d(x, s) & x \notin S \\ -\min_{s \notin S} d(x, s) & x \in S \end{cases}$$

- S = subset of pixels
- s = boundary
- d = Euclidean distance

# OCONet – step3 main




- Comparison with other representations

| | |
|---|---|
| L1 loss | • L1 encourages the model to output 1 or 0<br>• Sharp edges<br>• Fail on thin, ambiguous structures |
| Cross Entropy loss | • encourages the model, at each pixel, to output its estimate of the probability that that pixel is part of the mask. T<br>• Blurry mask |

# OCONet – step3 main

- Network loss functions

$$\mathcal{L}_{\text{object}} = \lambda_{\text{pixel}} \mathcal{L}_{\text{pixel}} + \lambda_{\text{mask}} \mathcal{L}_{\text{mask}}.$$

- where, $\lambda_{\text{pixel}} = 1$ and $\lambda_{\text{mask}} = 0.1$

- Mask loss: $$\mathcal{L}_m = \frac{1}{N_{pix}} \left\| M_{pred} - SDF(M_{GT}) \right\|_1$$

where, $N_{\text{pix}}$ is the number of pixels, $M_{GT}$ is the ground truth mask, and SDF is the signed distance function scaled to be between -1 and 1

- L2 reconstruction loss: $$\mathcal{L}_{\text{pixel}} = \frac{\left(\sum M_{\text{GT}}(x,y)(I_{\text{pred}}(x,y) - I_{\text{GT}}(x,y)\right)^2}{\sum M_{\text{GT}}(x,y)}$$

# OCONet – step3 Generator

- Generator loss functions

$$\mathcal{L}_{\text{gen}} = \lambda_{\text{adv}} \frac{1}{N_{\text{pix}}} \left( \sum_{(x,y)} -D(x,y) \right) + \mathcal{L}_{\text{object}} + \lambda_{\text{fm}} \mathcal{L}_{\text{fm}}$$

- where, $\lambda_{\text{perceptual}} = \lambda_{\text{pixel}} = 1 \text{ and } \lambda_{\text{mask}} = 0.1$

- Reconstruction loss: $\mathcal{L}_{object} = \frac{1}{N_{pix}} \left\| M_{pred} - SDF(M_{GT}) \right\|_1$

where $N_{\text{pix}}$ is the number of pixels, $M_{GT}$ is the ground truth mask, and SDF is the signed distance function scaled to be between -1 and 1

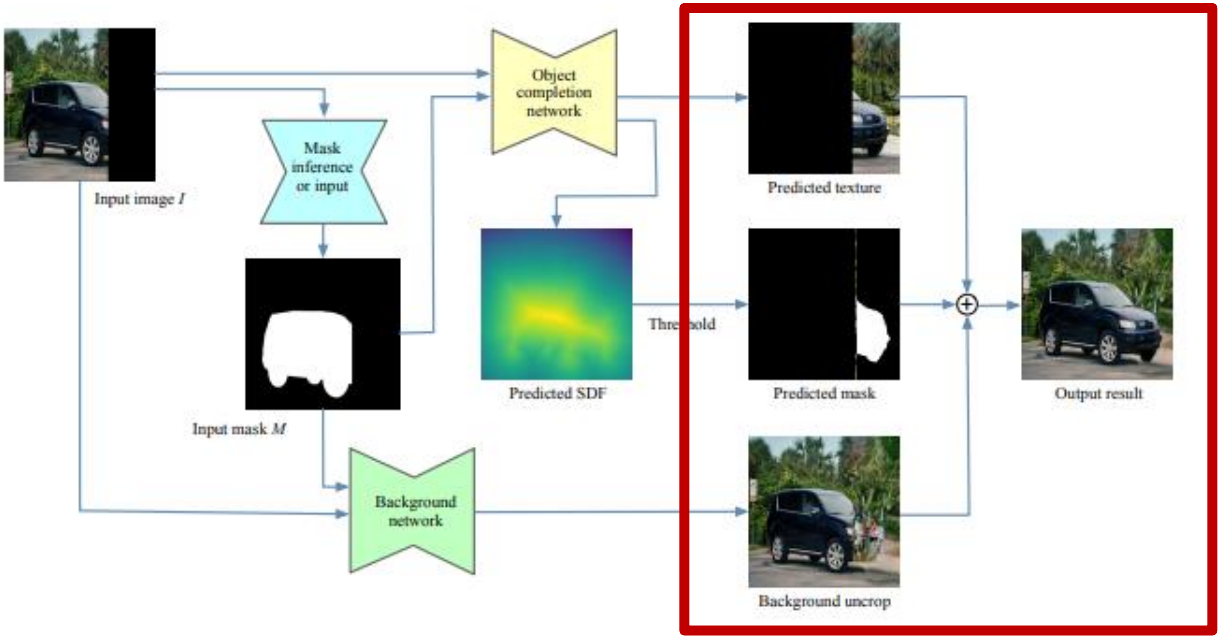
- Feature Matching loss: $\mathcal{L}_{\text{fm}} = \sum_i \frac{1}{N_i} \sum_{x,y} \left( \hat{\phi}_i(I_{\text{real}}) - \hat{\phi}_i(I_{\text{gen}}) \right)^2$

φi being features in the ith layer of the discriminator

- Gan loss: $\mathcal{L}_{GAN} = \frac{1}{N_{\text{pix}}} \left( \sum_{(x,y)} -D(x,y) \right)$

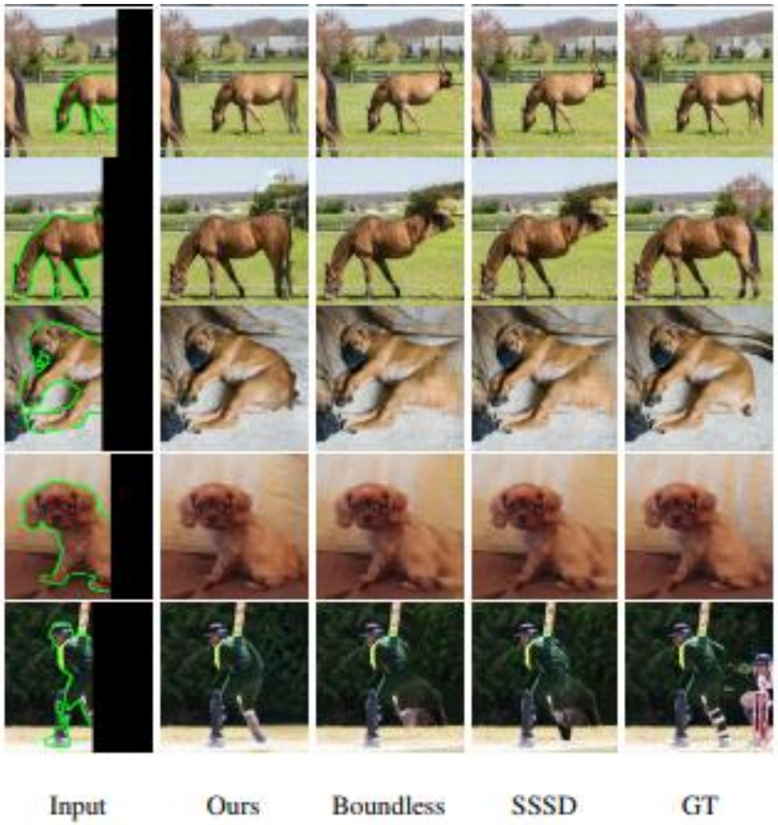
# OCONet – step4 composition

- composite foreground object onto an extrapolated background
  - Input: cropped color image I of size H*W*3
  - Output: mask indicating an object to be completed of size H*W*1
- Any instance segmentation method possible (in experiment shape mask)



# **OCONet** – experiment results

- Qualitative



# OCONet – experiment results

- Quantitative

| Category | $n$ | FID (lower is better) Ours | Boundless | SSSD | L1 (lower is better) Ours | Boundless | SSSD |
|---|---|---|---|---|---|---|---|
| Airplane | 1256 | **34.15** | 57.33 | 56.61 | **11.2** | 11.4 | 18.7 |
| Apple | 337 | **83.21** | 107.31 | 122.54 | **18.5** | **18.5** | 18.8 |
| Car | 1396 | **44.00** | 63.48 | 68.41 | **21.3** | 22.2 | 22.8 |
| Cat | 613 | **94.04** | 126.68 | 131.85 | **18.0** | 18.4 | 18.9 |
| Dog | 840 | **74.93** | 89.15 | 92.11 | **17.9** | 18.6 | 19.0 |
| Horse | 909 | **63.21** | 90.58 | 90.31 | **20.3** | 21.0 | 21.2 |
| Kite | 104 | **136.60** | 148.99 | 141.61 | 6.46 | **6.44** | 6.58 |
| Person | 802 | **107.36** | 112.08 | 112.46 | **19.8** | 20.2 | 20.4 |
| Train | 261 | **65.08** | 114.44 | 111.36 | **20.6** | 21.3 | 21.8 |
| All | 6518 | **20.82** | 30.67 | 32.02 | **17.9** | 18.4 | 18.8 |