IMAGE RECONSTRUCTION WITH INFORMED MASK GENERATION

ASEER • HARSH

BACKGROUND & MOTIVATION

- Current approaches solve image inpainting by passing an explicit mask
- However, those approaches are not scalable
  - Natural occurring Image Deformations do not come with a defined mask and are random in nature
  - Hand creating a mask for such images is a tedious process

OBJECTIVE

- To solve the problem of reconstruction by jointly performing inpainting and image enhancements.

- Our idea is concerned with reconstructing images using only the image as input.

OVERVIEW

1. Data
2. Network Architecture
3. Results
4. Conclusion

DATA : DATASETS

CelebFaces Attributes Dataset(CelebA)

- It is a public dataset and contains more that 200K celebrity faces
- Represents 10,000 unique identities with a large pose variations

Paris Street View

- This dataset contains street images for several cities
- It is composed of 15000 images
- The image's resolution is for 936 537 pixels.

DATA : MASKS

STILL WORKING …

OVERVIEW

1. Data
2. Network Architecture
3. Results
4. Conclusion

MASK PREDICTION NETWORK (MPN)

1. So we begin at this part of the network with input as the image.

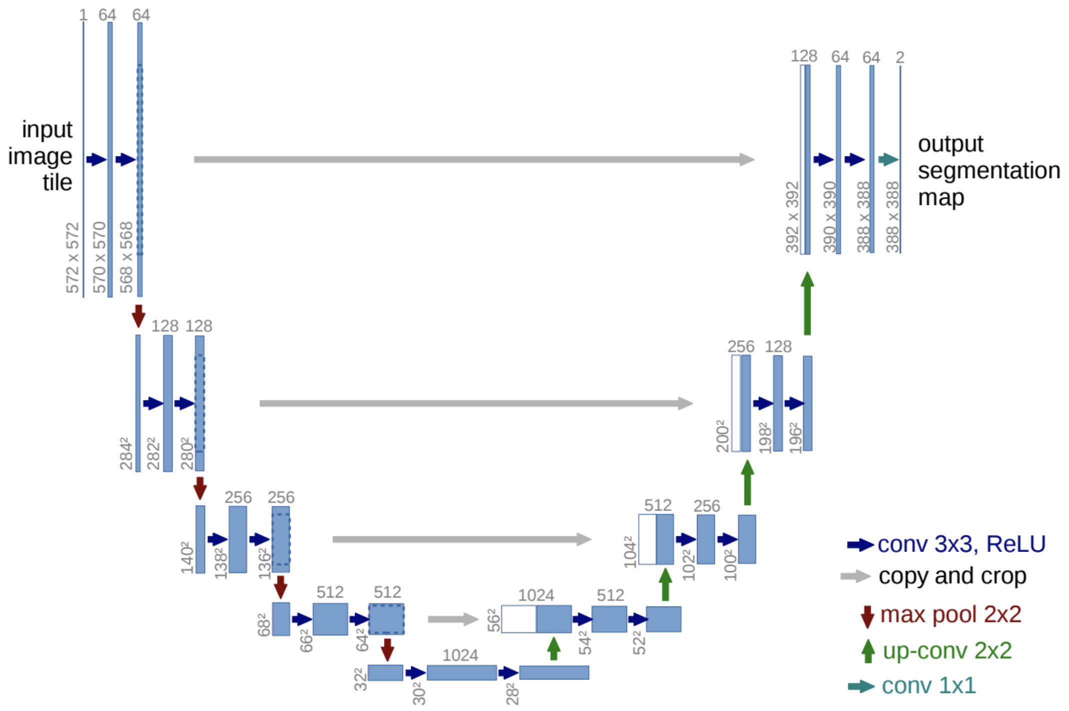
2. This network is based on the U-Net Segmentation Network. We modify the network to maintain the same input and output dimensions. The original network has the output dimensions smaller which worked for the original biomedical image segmentations requirement.

3. We began by experimenting on an Autoencoder architectures but due to the bottleneck of lower dimension layer and long length of the network learning eventually becomes difficult and results were poor.

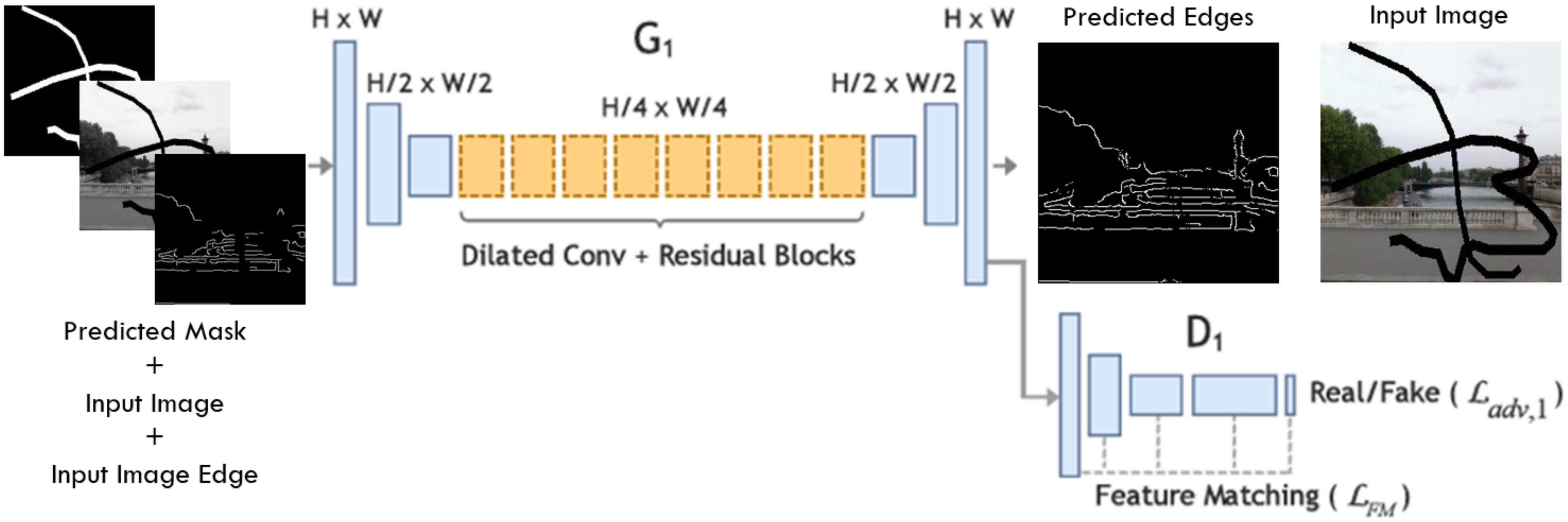
- We then used skip connections and 2D dropout and eventually used U-Net architecture for any further changes.

4. Loss

- Self-Adaptive BCE loss
- BCE loss



EDGE GENERATOR NETWORK



EDGE GENERATOR NETWORK

- We use the edge map generator module from EdgeConnect paper
  - Encoders down-sample twice, followed by eight residual blocks
  - Decoders upsample images back to the original size.
  - Dilated convolutions are used in the residual layers – to increase the receptive field
  - Spectral Normalization is applied on regular convolution for stable training
- For discriminators
  - We use a 70×70 PatchGAN architecture – determines if image patches of size 70×70 are real
- For the valid image pixels, we detect edges using Canny Edge Detector
- For the missing image pixels, we generate edges using the Generator network

IMAGE INPAINTING NETWORK

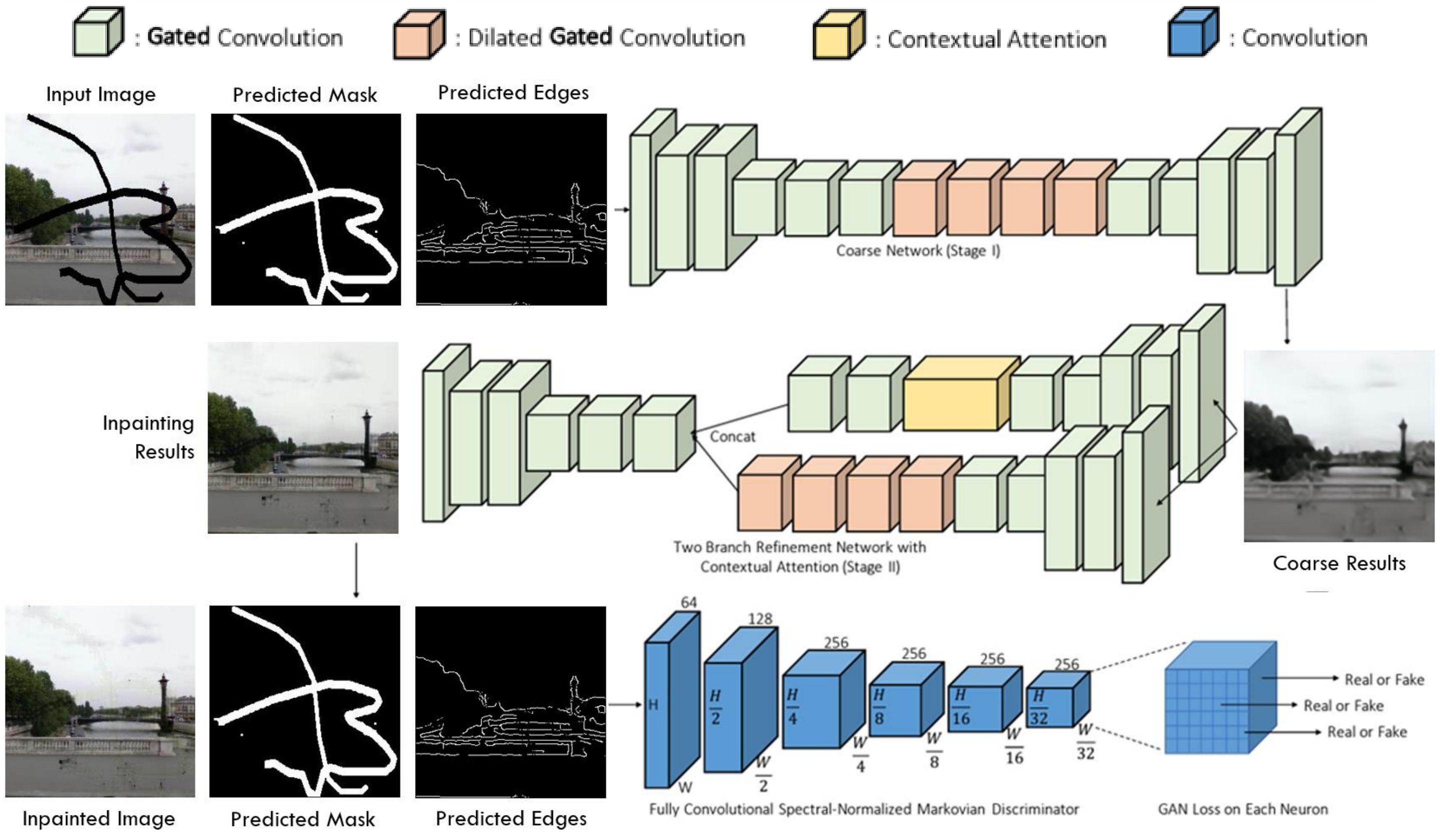


IMAGE INPAINTING GENERATOR NETWORK

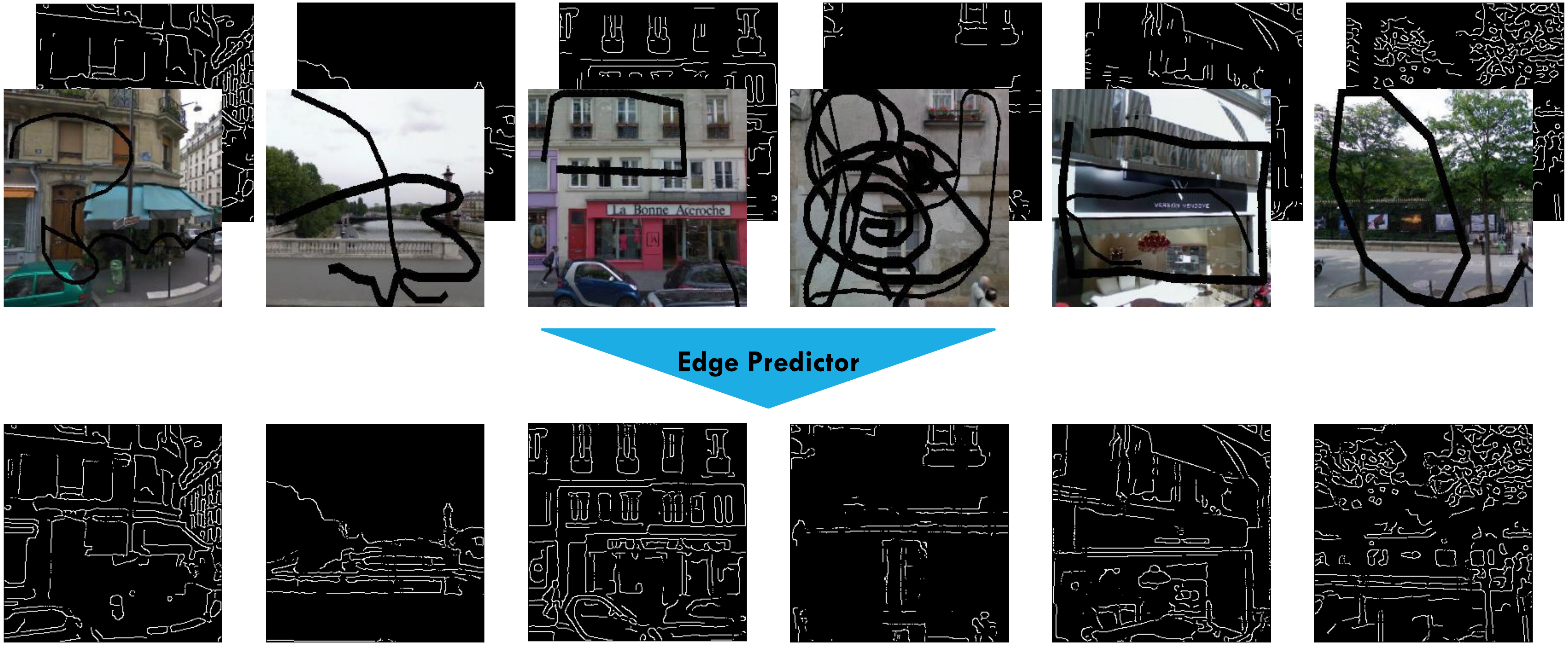
- Generator Network - takes the image , mask and edge map as input
  - Two step inpainting approach – First Coarse Network generates Coarse image, which is then passed through the Refinement Network to generate clear inpainted results
  - Coarse & Refinement generator – Use an encoder-decoder network of Gated Convolutions, without any skip connections
  - Refinement Network – additional branch with Contextual Attention network
- For Discriminator we use a fully convolutional network with spectral norm
  - Receptive field of each neuron in output covers entire input image – global discriminator is not necessary

OVERVIEW

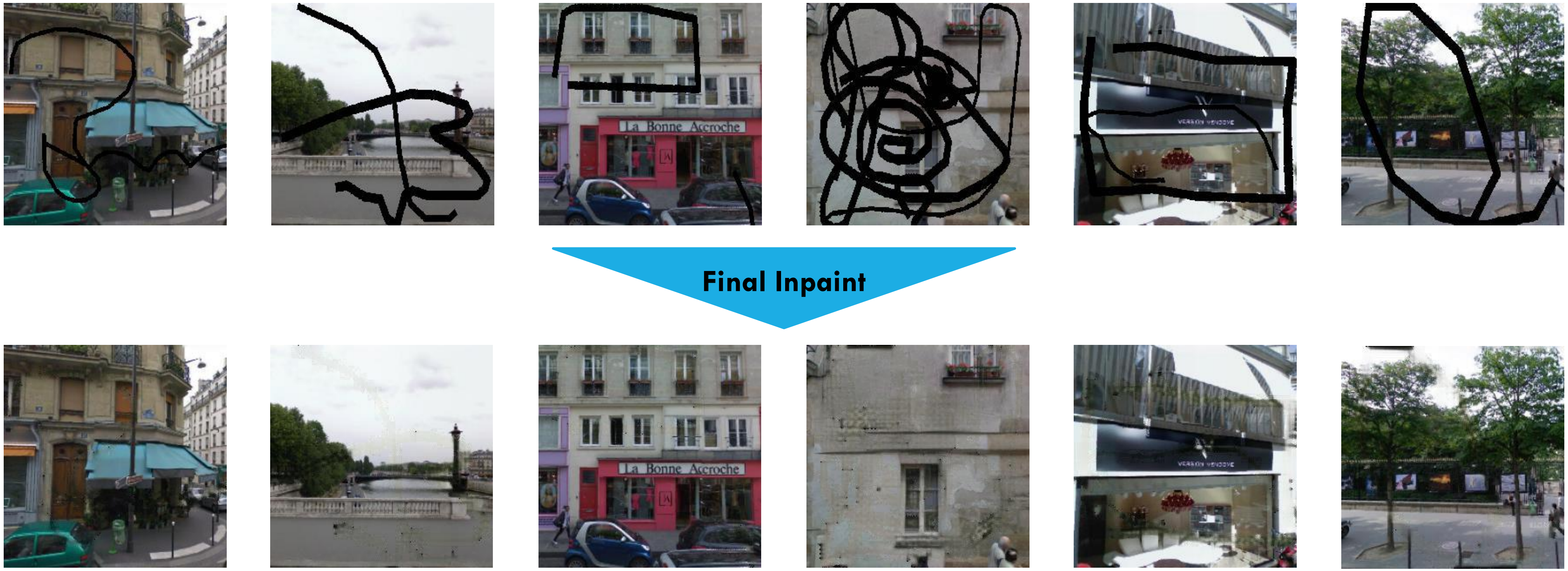
1. Data
2. Network Architecture
3. Results
4. Conclusion

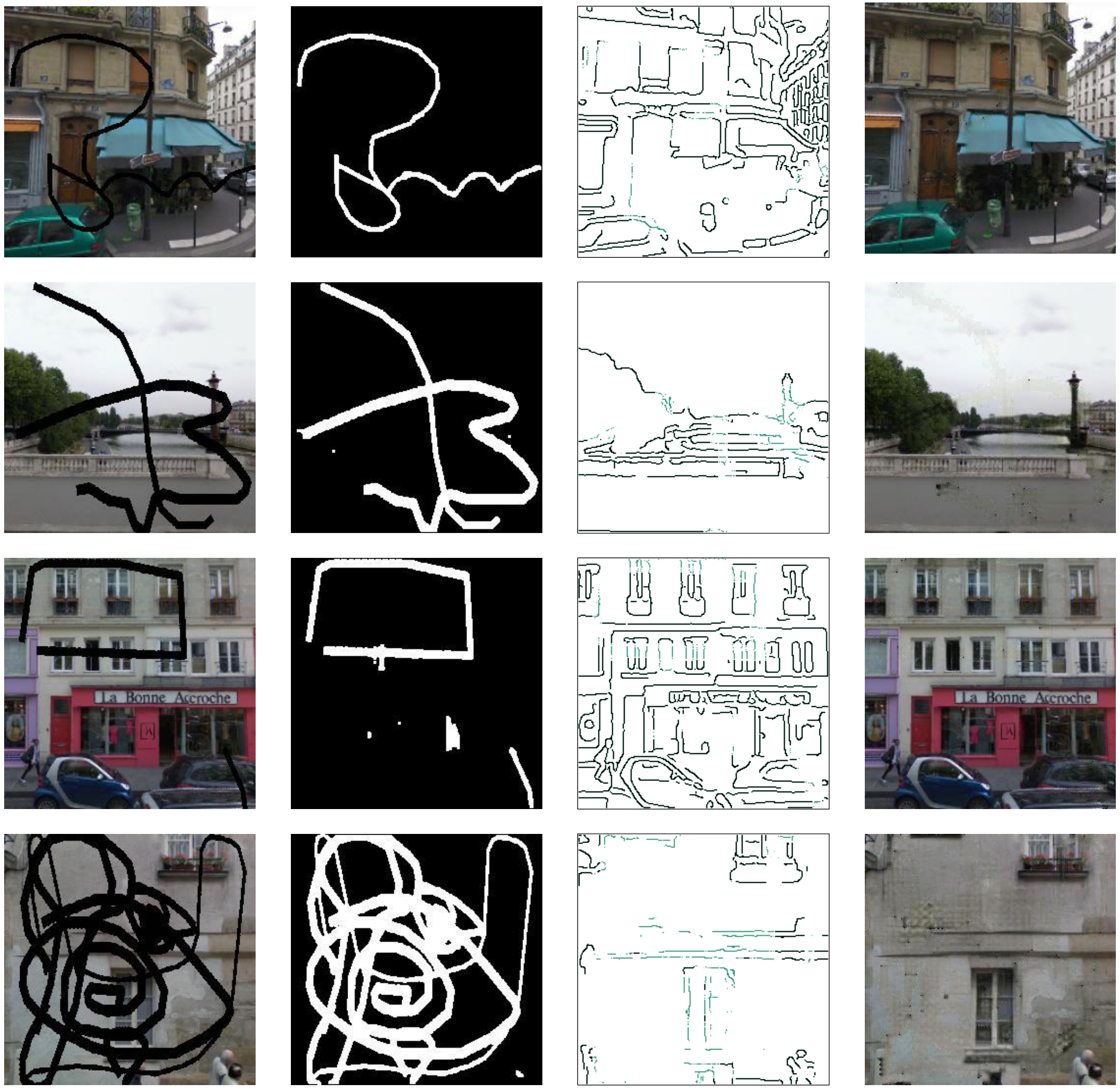
RESULTS – MASK PREDICTION NETWORK

RESULTS – EDGE GENERATOR NETWORK

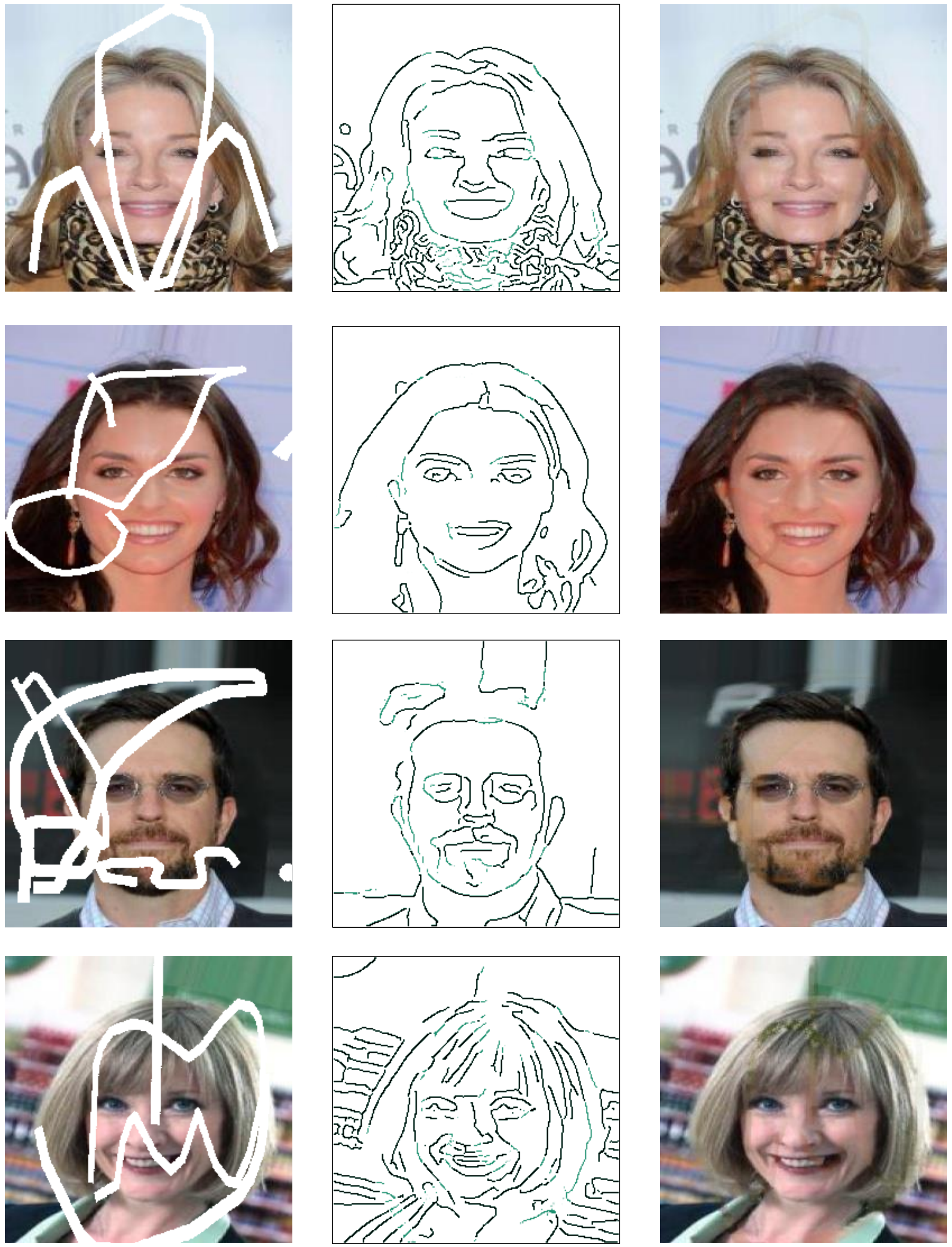


RESULTS – IMAGE INPAINTING NETWORK



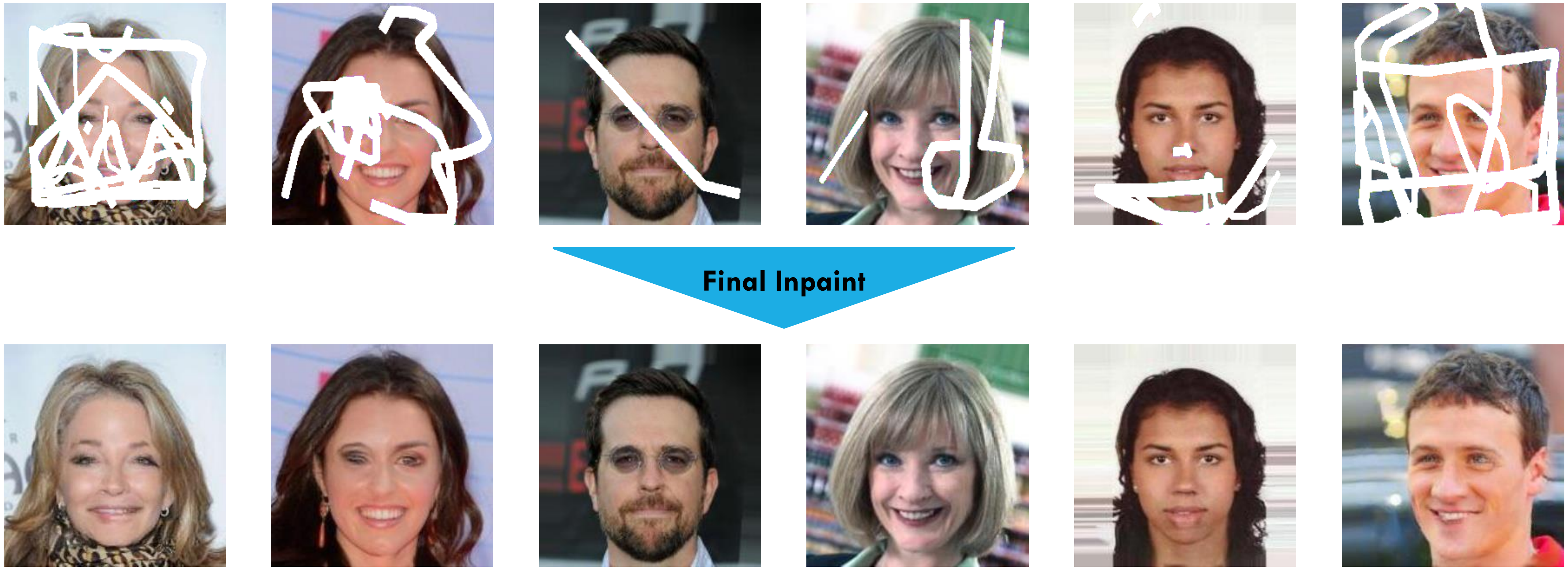


(a) (b) (c) (d)



(a) (b) (c)

RESULTS – IMAGE INPAINTING NETWORK



OVERVIEW

1. Data
2. Network Architecture
3. Results
4. Conclusion

CONCLUSIONS & FUTURE WORK

- We generate free form mask for input image
- Edges generated using Edge Predictor serve as exemplar guidance for inpainting
- Edges combined with Input Image and Mask help us obtain visually realistic images

Future Work

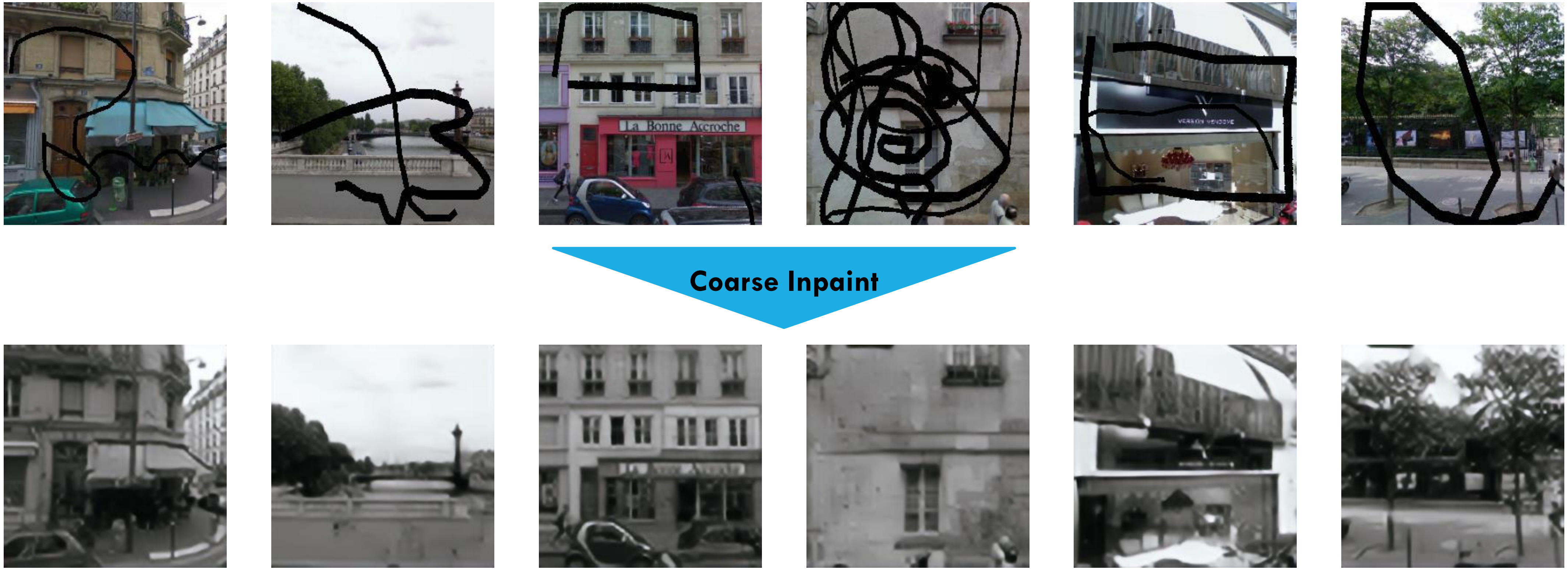
1. Joint training
2. Experiment with other loss types such as style loss, total variation loss etc.
3. Augment mask type to include multi-channel masks.

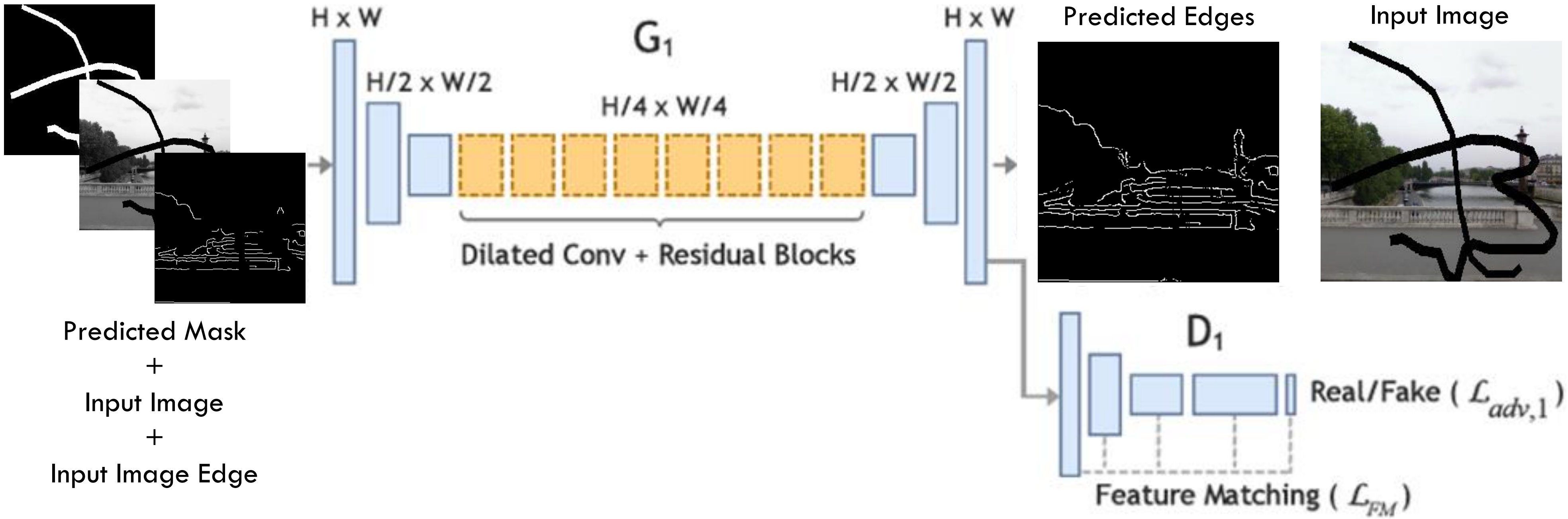
BIBLIOGRAPHY

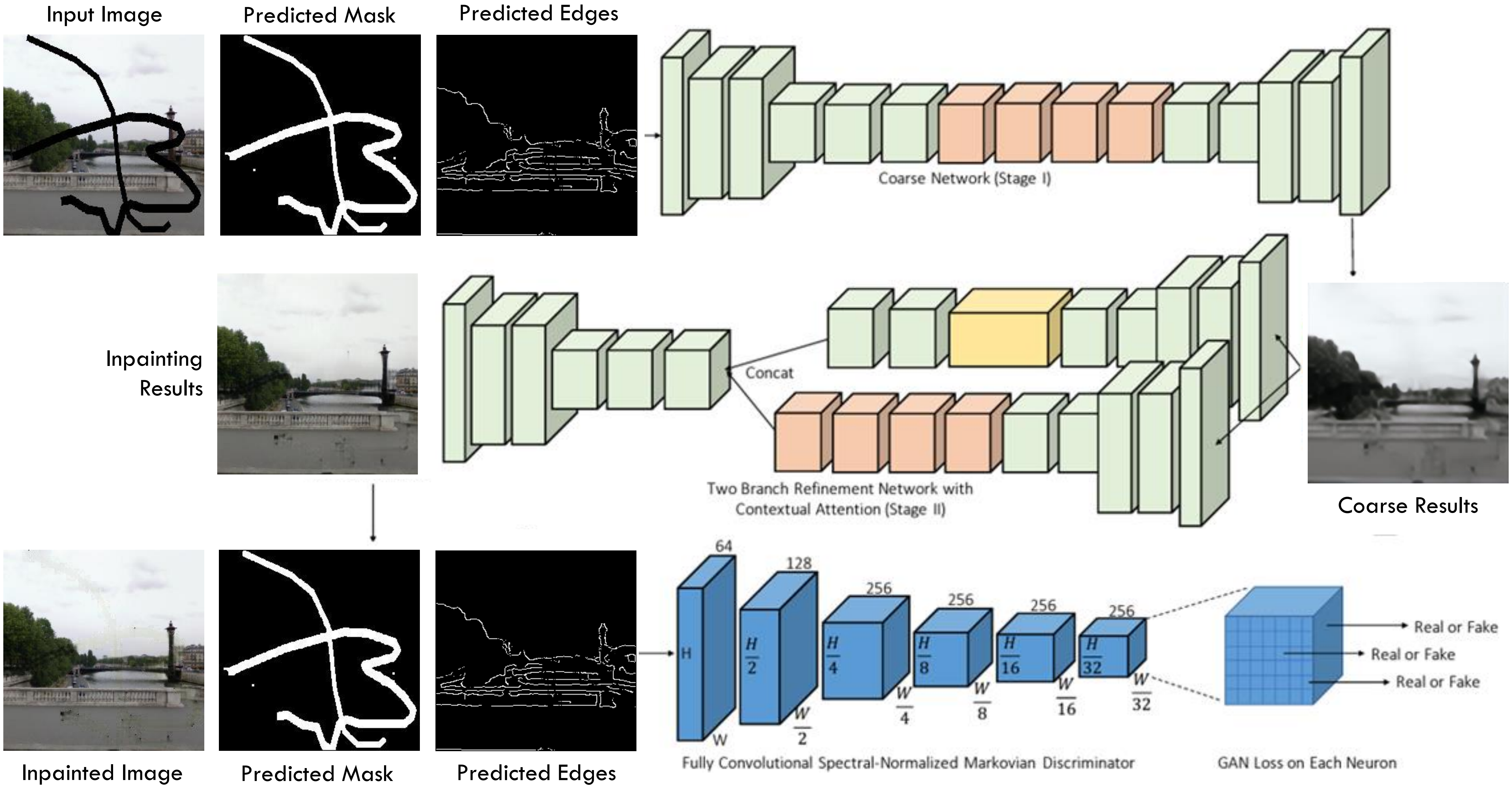
- Liu, Z., Luo, P., Wang, X., & Tang, X. (2018). Large-scale celebfaces attributes (celeba) dataset. Retrieved August, 15(2018), 11.
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2536-2544).
- Iskakov, K. (2018). Semi-parametric image inpainting. arXiv preprint arXiv:1807.02855.
- Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
- Nazeri, K., Ng, E., Joseph, T., Qureshi, F. Z., & Ebrahimi, M. (2019). Edgeconnect: Generative image inpainting with adversarial edge learning. arXiv preprint arXiv:1901.00212.
- Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2019). Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4471-4480).

ARCHIVE

RESULTS – IMAGE INPAINTING NETWORK




H x W
H/2 x W/2
G1
H/4 x W/4
H/2 x W/2
H x W
Predicted Edges
Input Image
Dilated Conv + Residual Blocks
D1
Real/Fake ($\mathcal{L}_{adv,1}$)
Feature Matching ($\mathcal{L}_{FM}$)
Predicted Mask
+
Input Image
+
Input Image Edge


Input Image
Predicted Mask
Predicted Edges
Coarse Network (Stage I)
Inpainting Results
Concat
Two Branch Refinement Network with Contextual Attention (Stage II)
Coarse Results
64
128
256
256
256
256
H
W
$\frac{H}{2}$ $\frac{W}{2}$
$\frac{H}{4}$ $\frac{W}{4}$
$\frac{H}{8}$ $\frac{W}{8}$
$\frac{H}{16}$ $\frac{W}{16}$
$\frac{H}{32}$ $\frac{W}{32}$
Real or Fake
Real or Fake
Real or Fake
Fully Convolutional Spectral-Normalized Markovian Discriminator
GAN Loss on Each Neuron
Inpainted Image
Predicted Mask
Predicted Edges