Face Image Inpainting using GANs

- Bhavana Talluri
- Sai Jahnavi Shanvitha Pasumarthy
- Lavanya Konda

Guide: Dr. Himangshu Sarma

Overview

- Problem statement
- Applications of Inpainting
- Traditional methods and Drawbacks
- Motivation
- Proposed approach
- Dataset description
- Training
- Results
- Future work

Problem statement - Face Image Inpainting

- Face image inpainting (or face completion) is the task of generating plausible facial structures for missing pixels in a face image.

- Aim: To recover or fill in missing information in occluded images. The goal is to produce more legible and visually realistic face images from an image with a masked region that has missing content.

Input

Result

Groundtruth

Image source : Predicted by our model

Applications

- Image restoration or completion
- Removal or replacement of selected objects
- Completing corrupted video frames due to over-compression

Traditional method

- Navier-Stokes method:
  - Introduces smoothness priors via Partial differential equations
  - Models image inpainting as a third-order Partial differential equations
  - To estimate missing pixels, take normalized weighted sum of pixels from a neighbourhood of the pixels.
  - To estimate color of the pixels, gradient of the neighbourhood pixels are used.

source : Bertalmio, Marcelo, Andrea L. Bertozzi, and Guillermo Sapiro. "Navier-stokes, fluid dynamics, and image and video inpainting." *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. IEEE, 2001.

Traditional method - drawback

- Navier-Stokes method:
  - Well suited for filling geometrical shapes and completing small regions in the image.
  - Unable to recover the texture of large areas, which tend to blur in these situations.

source : Yuan, Liuchun, Congcong Ruan, Haifeng Hu, and Dihu Chen. "Image inpainting based on patch-GANs." *IEEE Access* 7 (2019): 46411-46421.

Traditional method

- Texture synthesis by non-parametric sampling:
  - Common idea: missing areas can be learned from similar regions in a sample
  - Some variations include using a quickly approximating nearest neighbour patch search algorithm and employ the best matching patches for reconstruction.
  - Other variations include combining the copy-and-paste texture synthesis, geometric partial differential equation and coherence neighbouring pixel to obtain better image inpainting result.

source : A.A.EfrosandT.K.Leung,“Texturesynthesisbynon-parametricsam- pling,” in *Proc. ICCV*, Sep. 1999, pp. 1033–1038.

Traditional method - drawback

- Texture synthesis by non-parametric sampling:
  - Works well for simple task like background completion, such as sky (implying similar pattern of the missing part should be contained in the existing regions)
  - These methods become invalid if the missing region is vast.

source : Yuan, Liuchun, Congcong Ruan, Haifeng Hu, and Dihu Chen. "Image inpainting based on patch-GANs." *IEEE Access* 7 (2019): 46411-46421.

Motivation

- Conditional GANs for inpainting:
  - Applying high-level features (extracted using CNN) to reconstruct damaged images based on the deep convolution networks.
  - cGANs have some conditional settings and they learn the image-to-image mapping under this condition, whereas basic GANs generate images from a random distribution vector with no condition applied.
  - This method proved to be more effective than prior methods.

Proposed Approach

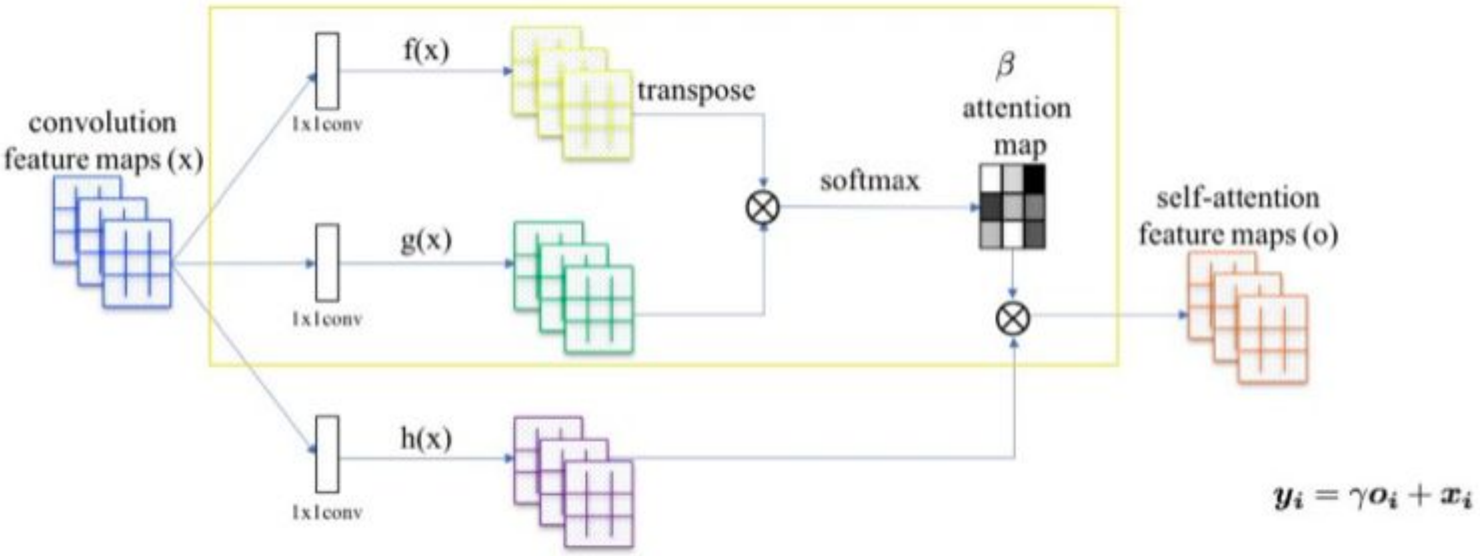
- Pix2Pix GAN (variation of CGAN) + Attention mechanism
  - Pix2Pix GAN:
    - Produces an output classification that classifies multiple patches in the input image pairs (patchGAN), producing an output of NxN dimension
    - Has generator and discriminator just like a normal GAN, but it's more supervised than GAN as it has target images as output labels.

Proposed Approach

- Pix2Pix GAN (variation of CGAN) + Attention mechanism
  - Attention mechanism:
    - For GAN models trained with ImageNet, they are good at classes with a lot of texture (landscape, sky) but perform much worse for structure.
    - While convolutional filters are good at exploring spatial locality information, receptive fields may not be large enough to cover larger structures.
    - Filter size or depth of deep network can be increased, but training gets harder. Alternatively, attention concept can be applied

Proposed Approach

- Attention mechanism:



source : Zhang, Han, et al. "Self-attention generative adversarial networks." *International Conference on Machine Learning*. 2019.

Proposed Approach

- Attention mechanism:

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{W_g}\boldsymbol{x} \qquad \boldsymbol{x} \in \mathbb{R}^{C \times N},\ \boldsymbol{W_f} \in \mathbb{R}^{\bar{C} \times C},\ \bar{C} = C/8$$

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{W_f}\boldsymbol{x} \qquad \boldsymbol{W_g} \in \mathbb{R}^{\bar{C} \times C}$$

$$s_{ij} = \boldsymbol{f}(\boldsymbol{x_i})^T \boldsymbol{g}(\boldsymbol{x_j})$$

$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})} \qquad \beta \in \mathbb{R}^{N \times N}$$

$\beta_{j,i}$ indicates the extent to which the model attends to the i^{th} location when synthesizing the j^{th} region.

source : Zhang, Han, et al. "Self-attention generative adversarial networks." *International Conference on Machine Learning*. 2019.

Proposed Approach

- Attention mechanism:

$$\boldsymbol{h}(\boldsymbol{x_i}) = \boldsymbol{W_h x_i} \qquad \boldsymbol{W_h} \in \mathbb{R}^{C \times C}$$

$$\boldsymbol{o_j} = \sum_{i=1}^{N} \beta_{j,i} \boldsymbol{h}(\boldsymbol{x_i}) \qquad \boldsymbol{o} \in \mathbb{R}^{C \times N}$$

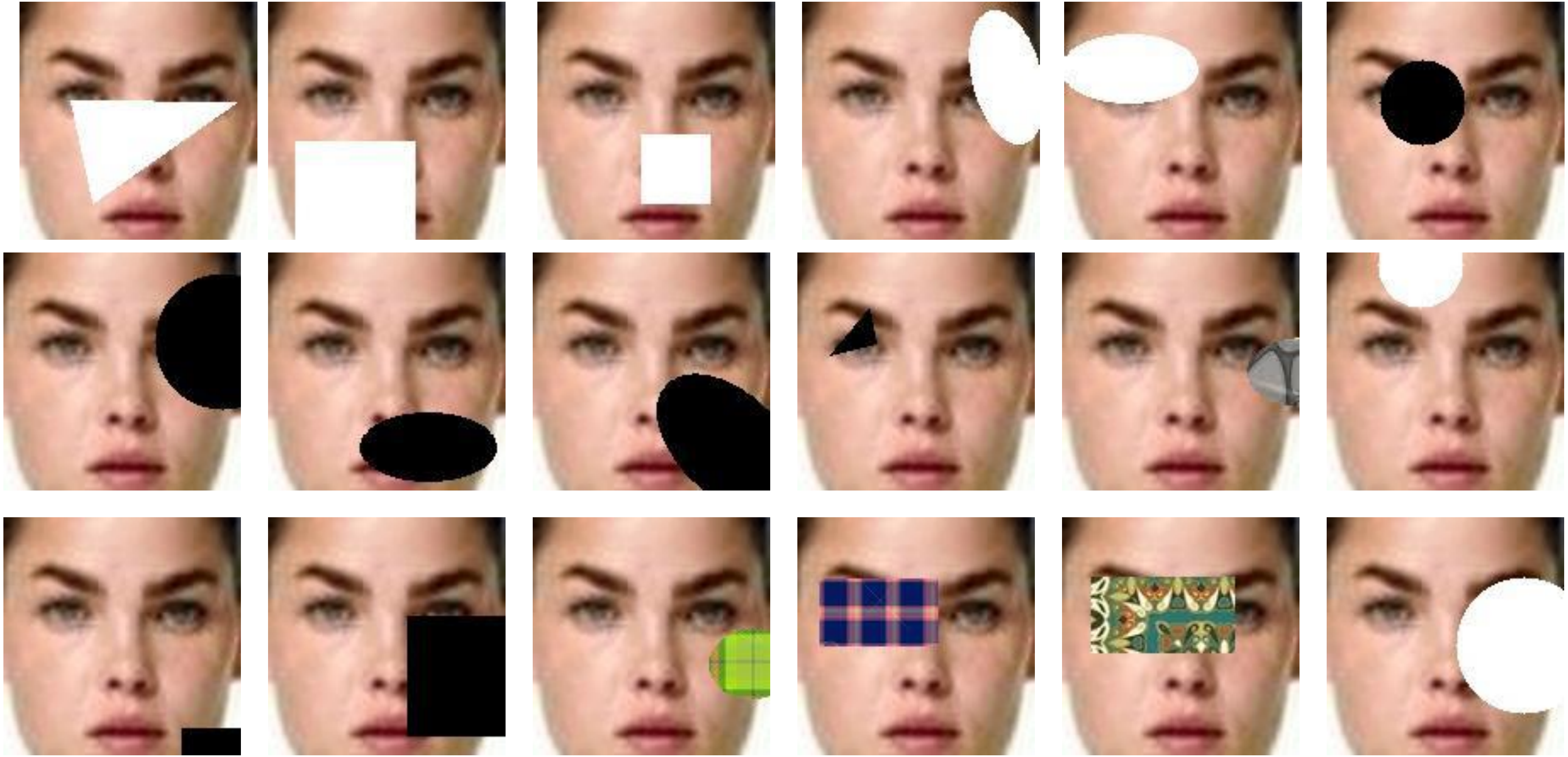
Final output of convolution layer: $\boldsymbol{y_i} = \gamma \boldsymbol{o_i} + \boldsymbol{x_i}$

source : Zhang, Han, et al. "Self-attention generative adversarial networks." *International Conference on Machine Learning*. 2019.

Dataset Description

- Dataset chosen: CelebA (CelebFace Attributes Dataset)
- Consists of 2,02,599 face images and included large pose variations and clutter background.
- We applied MTCNN to tightly crop the face images to get accurate results.
- We then wrote a python script to generate patches of different shapes, sizes and textures to be superimposed on the images in CelebA dataset.

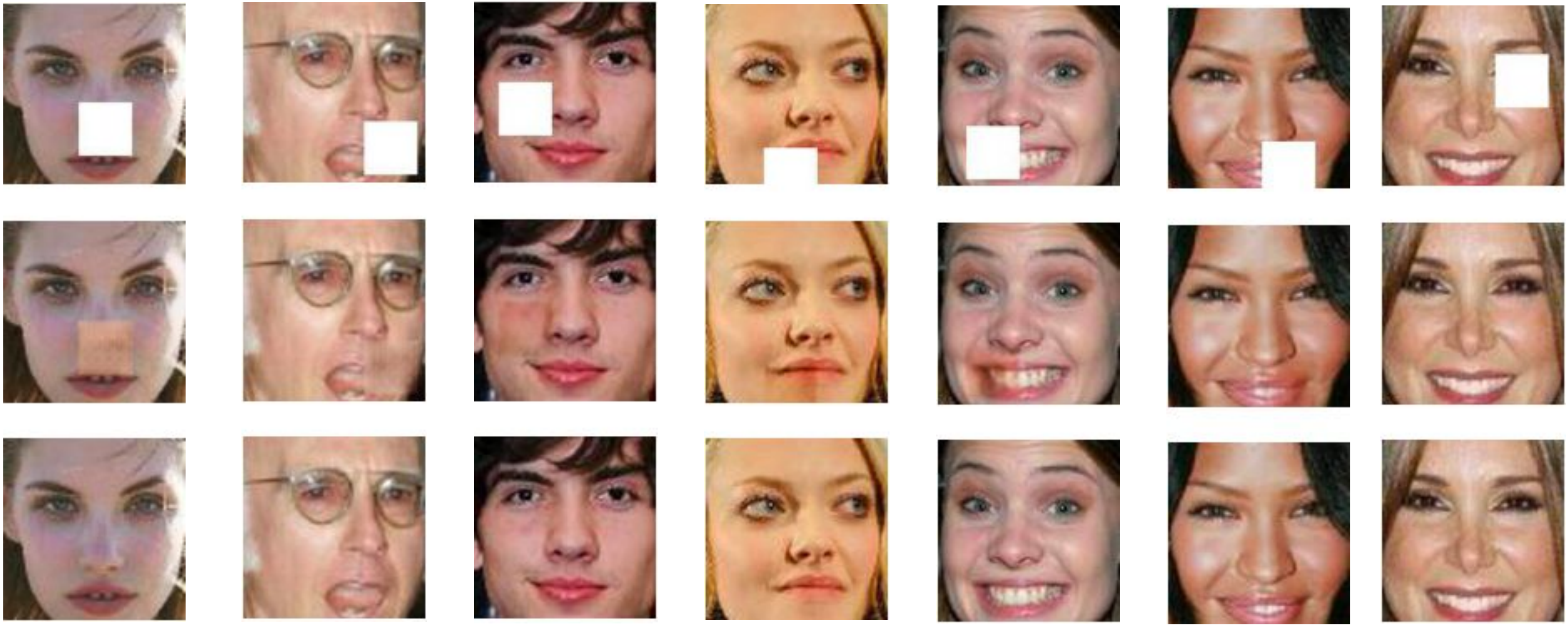
source : Liu, Ziwei, et al. "Large-scale celebfaces attributes (celeba) dataset." *Retrieved August* 15 (2018): 2018.



Training

- Model: Pix2Pix GAN + attention mechanism
- Training images: 1098 (white square only)
- Batch_size: 1
- Epochs: 100
- Total training steps: 109800
- GPU: GeForce GTX 1080

Results



10 epochs | 30 epochs | 40 epochs | 60 epochs | 80 epochs | 90 epochs | 100 epochs

Results

- **PSNR:** computes the "peak signal-to-noise ratio", in decibels between two images. Higher the PSNR, the better is the similarity.

- **SSIM:** "Structural Similarity Index" is a method for measuring similarity between two images. Higher the SSIM, better is the similarity.

Model	Dataset	PSNR	SSIM
Patch GANs (IEEE)	CelebA	23.07	0.923
Ours	CelebA	21.567	0.889

Future Work

- Currently the experimentation has only been done using images with white square patches only. We would further like to train the model on the entire dataset to obtain a robust model.

- Currently, the training process for Cycle GAN with attention mechanism is under process, we would like to perform similar analogy on this model too once the training is completed.

- We also want to explore how other Generative Adversarial Networks with attention mechanism will perform on the CelebA dataset

References

Yuan, Liuchun, Congcong Ruan, Haifeng Hu, and Dihu Chen. "Image inpainting based on patch-GANs." *IEEE Access* 7 (2019): 46411-46421.

Zhang, Han, et al. "Self-attention generative adversarial networks." *International Conference on Machine Learning*. 2019.

Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Bertalmio, Marcelo, Andrea L. Bertozzi, and Guillermo Sapiro. "Navier-stokes, fluid dynamics, and image and video inpainting." *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. IEEE, 2001.

A.A.EfrosandT.K.Leung,‘‘Texturesynthesisbynon-parametricsam- pling,’’ in *Proc. ICCV*, Sep. 1999, pp. 1033–1038.

Thank you!!