Generative Image Inpainting for Person Pose Generation

Anubha Pandey, Vismay Patel

Indian Institute of Technology Madras

cs16s023@cse.iitm.ac.in

19th September 2018

Overview

Problem Statement

Chalearn LAP Inpainting Competition Track1 - Inpainting of still images of humans

- **Objective** To restore the masked parts of the image in a way that resembles the original content and looks plausible to a human.

Problem Statement

Chalearn LAP Inpainting Competition Track1 - Inpainting of still images of humans

- **Objective** To restore the masked parts of the image in a way that resembles the original content and looks plausible to a human.
- **Dataset**
  - The dataset consists of images with multiple square blocks of black pixels randomly placed, occluding at most 70% of the original image.
  - The dataset is taken from multiple sources- MPII Human Pose Detection, Leeds Sports Pose Dataset, Synchronic Activities Stickmen V, Short BBC Pose and Frames labelled in Cinema.
  - 28755 training samples, 6160 validation samples and 6160 test samples.

Introduction

- Image Inpainting is the task of filling missing pixels of an image.

Introduction

- Image Inpainting is the task of filling missing pixels of an image.
- The main challenge of the task is to generate realistic and semantically plausible pixel for the missing regions that blends properly with the existing image pixels.

Related Works

- Early works [1] [2] [3] use patch based methods to solve the problem.
  - They copy matching background patches into the holes.
  - These paper works well in background inpainting tasks.
  - They can't synthesize novel structures.

Related Works

- New deep methods use CNN and GAN networks to formulate the solution and have produces promising results for image inpainting.

Related Works

- New deep methods use CNN and GAN networks to formulate the solution and have produces promising results for image inpainting.
  - These methods train encoder-decoder network jointly with adversarial networks to produce pixels which are coherent with the existing ones.

Related Works

- New deep methods use CNN and GAN networks to formulate the solution and have produces promising results for image inpainting.
    - These methods train encoder-decoder network jointly with adversarial networks to produce pixels which are coherent with the existing ones.
    - They can't model long term correlations between distant contextual information and hole regions.

Related Works

- New deep methods use CNN and GAN networks to formulate the solution and have produces promising results for image inpainting.
  - These methods train encoder-decoder network jointly with adversarial networks to produce pixels which are coherent with the existing ones.
  - They can't model long term correlations between distant contextual information and hole regions.
  - Produces boundary artifacts, distorted structures, blurry textures inconsistent with surroundings.

Related Works

- More recently, Globally and locally consistent image completion [4] CVPR 2017 paper, improve the results by introducing local and global discriminators. In addition, it uses dilated convolutions to increase the receptive fields and replace the fully connected layers adopted in the contextual encoders.

Proposed Solution: Image Inpainting generator with skip-connections

- We use an encoder-decoder based CNN model with a combination of regular and dilated convolutions followed by batch normalization and ReLU to encode the partial image.

Proposed Solution: Image Inpainting generator with skip-connections

- We use an encoder-decoder based CNN model with a combination of regular and dilated convolutions followed by batch normalization and ReLU to encode the partial image.
- The decoder uses skip connections from the encoder and combination of deconvolution and convolutions to generate the full image.

Proposed Solution: Image Inpainting generator with skip-connections

- We use an encoder-decoder based CNN model with a combination of regular and dilated convolutions followed by batch normalization and ReLU to encode the partial image.
- The decoder uses skip connections from the encoder and combination of deconvolution and convolutions to generate the full image.
- **Inputs**
  - The input to the model is 128*128*4 sized tensor which is concatenation of the input image and the mask.
  - We use data available in the 'maskdata.json' file to generate binary mask images. The masks contain ones in places of holes and zeros everywhere else.

Network Architecture

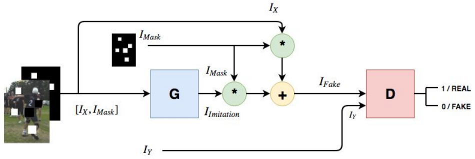

Figure: Architecture of the discriminator module of the inpainting network. Each building block is described in Figure 9.

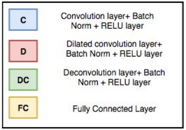

Figure: Building blocks of the network.

Network Architecture

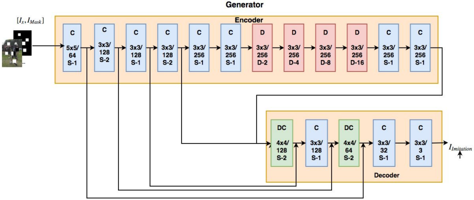

Figure: Architecture of the generator module of the inpainting network. Building block is shown in Fig 9.

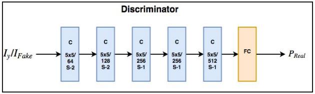

Figure: Architecture of the discriminator module of the inpainting network.Building block is shown in Fig 9.

Loss Functions

Following loss functions have been used to train the network-

- **Reconstruction Loss** [5]
  $L_r = \frac{1}{K}\sum_{i=1}^{K}|I_x^i - I_{imitation}^i| + \alpha * \frac{1}{K}\sum_{i=1}^{K}(I_{Mask}^i - I_{Mask}^i)^2$
  where, K is the batch size and alpha = 0.000001 and $I_{imitation}^i$ is the output of the decoder.

Loss Functions

Following loss functions have been used to train the network-

- **Reconstruction Loss** [5]
  $L_r = \frac{1}{K}\sum_{i=1}^{K}|I_x^i - I_{imitation}^i| + \alpha * \frac{1}{K}\sum_{i=1}^{K}(I_{Mask}^i - I_{Mask}^i)^2$
  where, K is the batch size and alpha = 0.000001 and $I_{imitation}^i$ is the output of the decoder.
- **Adversarial Loss** [5] $L_{real} = -log(p)$, $L_{fake} = -log(1-p)$
  $L_d = L_{real} + \beta * L_{fake}$
  where, p is the output probability of the discriminator module and β = 0.01 (hyper parameter)

Loss Functions

Following loss functions have been used to train the network-

- **Reconstruction Loss** [5]
  $L_r = \frac{1}{K}\sum_{i=1}^{K}|I_x^i - I_{imitation}^i| + \alpha * \frac{1}{K}\sum_{i=1}^{K}(I_{Mask}^i - I_{Mask}^i)^2$
  where, K is the batch size and alpha = 0.000001 and $I_{imitation}^i$ is the output of the decoder.
- **Adversarial Loss** [5] $L_{real} = -log(p)$, $L_{fake} = -log(1-p)$
  $L_d = L_{real} + \beta * L_{fake}$
  where, p is the output probability of the discriminator module and β = 0.01 (hyper parameter)
- **Perceptual Loss** [6]
  $L_p = \frac{1}{K}\sum_{i=1}^{K}(\phi(I_y) - \phi(I_{imitation}))^2$
  where, ϕ represents features from VGG16 network pretrained on Microsoft COCO dataset.

Training

- The network is trained using Adam Optimizer with learning rate 0.001 and batch size 12.

Training

- The network is trained using Adam Optimizer with learning rate 0.001 and batch size 12.
- For first 5 epochs only the generator module of the network is trained minimizing only the reconstruction loss and perceptual loss

Training

- The network is trained using Adam Optimizer with learning rate 0.001 and batch size 12.
- For first 5 epochs only the generator module of the network is trained minimizing only the reconstruction loss and perceptual loss
- For the next 15 epochs, the entire GAN network [5] is trained end-to-end minimizing Adversarial and Perceptual loss.

Results

- With our proposed solution we secured **2nd position** in the competition.

Results

- With our proposed solution we secured **2nd position** in the competition.
- To evaluate the quality of the reconstruction, metrics as mentioned on the competition's website are used.

Evaluation Metrics	Training Phase	Testing Phase
PSNR	20.4314	21.5118
MSE	0.0176	0.0158
DSSIM	0.2089	0.2048
WNJD	0.1488	0.1495

Results

Figure: Input Image

Figure: Generated Image

Figure: Ground Truth

Results

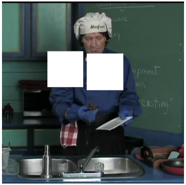

Figure: Input Image

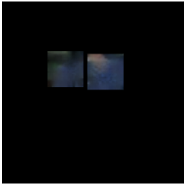

Figure: Generated Image

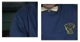

Figure: Ground Truth

Results

Figure: Input Image

Figure: Generated Image

Figure: Ground Truth

Conclusion

- We propose a generative solution for the image inpainting task.

Conclusion

- We propose a generative solution for the image inpainting task.
- We have trained our model to generate patches which has not appear anywhere in the scene.

Conclusion

- We propose a generative solution for the image inpainting task.
- We have trained our model to generate patches which has not appear anywhere in the scene.
- Also, it has learn to inpaint images with randomly placed masks of variable size.

Future Work

- We aim to improve the resolution of the inpainted image by using multi-stage GANs at different resolution.

Future Work

- We aim to improve the resolution of the inpainted image by using multi-stage GANs at different resolution.
- Moreover, techniques to handle the multiple modalities of the image and using loss functions related to pose estimation would help improve the results.

Thank You

References

Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang.
Generative image inpainting with contextual attention.
arXiv preprint, 2018.

Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman.
Patchmatch: A randomized correspondence algorithm for structural image editing.
ACM Transactions on Graphics (ToG), 28(3):24, 2009.

James Hays and Alexei A Efros.
Scene completion using millions of photographs.
In *ACM Transactions on Graphics (TOG)*, volume 26, page 4. ACM, 2007.

Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa.
Globally and locally consistent image completion.
ACM Transactions on Graphics (TOG), 36(4):107, 2017.