Title of the contribution

July 20, 2018

1 Team details

- Team name : UNLU
- Team leader name : Gizem Esra Ünlü
- Team leader address, phone number and email :
  Address: Halıcıoğlu Mahallesi, Erenler Tekkesi Sokak, Selam Apartmani, No:5, Daire:5, Beyoğlu/İstanbul Türkiye
  Phone Number: +90 542 720 53 80
  e-mail: unlugi@hotmail.com
- Rest of the team members: None
- Team website URL (if any): None
- Affiliation: Bogazici University - Turkey

2 Contribution details

- Title of the contribution
  People Inpainting with Generative Adversarial Networks
- Final score
  PSNR: 21.8711893588
  MSE: 0.0158471260207
  DSSIM: 0.208834181594
  WNJD: 0.148852195872
- General method description
  In this work, a GAN-like architecture (as in [2]) is used with three subnetworks: generator, local discriminator, global discriminator. The generator is a feed-forward CNN that fills the masked sections of the image using dilated convolutional layers. The believablity of the inpainted sections of the image is enabled by the local and global discriminators which ensure the resulting images are consistent with the rest of the image. The

main contribution of this solution was achieved by implementing a new loss function, Generalized Loss-Sensitive GAN, which is a generalized version of the Wasserstein loss used in the original implementation, for better inpainting results.

- References
  1) Iizuka, Satoshi, Edgar Simo-Serra, and Hiroshi Ishikawa. "Globally and locally consistent image completion." ACM Transactions on Graphics (TOG) 36.4 (2017): 107.

  2) Yu, Jiahui, et al. "Generative image inpainting with contextual attention." arXiv preprint (2018).

  3) Qi, Guo-Jun. "Loss-sensitive generative adversarial networks on lipschitz densities." arXiv preprint arXiv:1701.06264 (2017).

- Representative image / diagram of the method

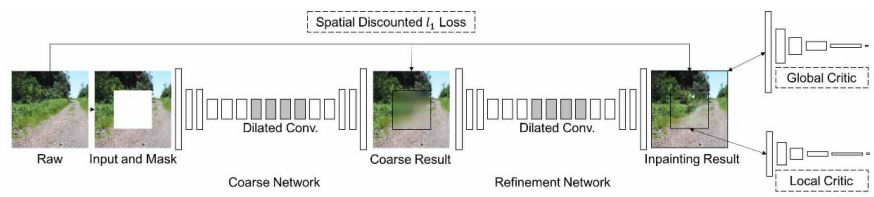

Figure 1: The architecture used in the solution. (Image reprinted from https://arxiv.org/pdf/1801.07892.pdf)

- Describe data preprocessing techniques applied (if any)
  Firstly, all the masked images and their corresponding mask labels are combined to create original images for training/validation sets for training the feedforward GAN architecture. During training all the input images are resized to 256x256 (the smaller side of the image is resized to 256 (the image becomes 256x?) then a 256x256 crop is taken randomly)

3 Method description

3.1 Features / Data representation

None

3.1.1 Dimensionality reduction

None

3.1.2 Compositional model

None

3.1.3 Learning strategy

None

3.1.4 Other techniques

None

3.1.5 Method complexity

Estimated method complexity

3.2 Data Fusion Strategies

None

3.3 Global Method Description

- Which pre-trained or external methods have been used (for any stage, if any)
  The solution is built on the architecture of the following paper 'Generative Image Inpainting with Contextual Attention'. This model's pretrained model on the Places2 dataset was used to finetune a new model with the challenge dataset.

- Which additional data has been used in addition to the provided ChaLearn training and validation data (at any stage, if any)
  None

- Qualitative advantages of the proposed solution
  The Generalize Loss Sensitive loss function applied to the inpainting architecture to obtain better looking results.

- Results of the comparison to other approaches (if any)
  None

- Novelty degree of the solution and if is has been previously published
  The inpainting architecture in this solution was taken wholly from the 'Generative Image Inpainting with Contextual Attention' paper. Generalized Loss-Sensitive GAN loss function was applied on top of this implementation. Finetuning was done with the challenge dataset on top of a pre-trained model (Places2 dataset) and parameter tuning was done.

4 Other details

- Language and implementation details (including platform, memory, parallelization requirements)
  Linux
  Python 3.6 - Tensorflow 18.01
  NVIDIA DGX-1 Server with Tesla V100 GPU with 16 GB memory (1 GPU was used during both training and testing)

- Human effort required for implementation, training and validation?
  4 weeks

- Training/testing expended time?
  Training: 50 hours, 10 Epochs (for 1 model)
  Testing: 12 hours(for 1 model)

- General comments and impressions of the challenge? what do you expect from a new challenge in face and looking at people analysis?

  The most challenging part of the competition for me was the dataset. First of all, the dataset came masked; the original images were not included, you had to combine the masked images and mask labels (obtaining original images) to train deep architectures. Also, all the images have different resolutions; this makes training CNN/GAN networks really hard since they accept fixed size input images during training and testing. However, I realize that inpainting in the wild is an important challenge therefore, in the future two types of challenges may be held with fixed size datasets and variable sized datasets or the submissions may be evaluated with respect to their performances in such datasets.

- Parameters used in the model in CODALAB
  WGAN-GP-LAMBDA: 5
  GLS-GAMMA: 0.001
  GLS-SLOPE: 0.5