CS4243 Project Plan (Group 4)

Chen Yuanbo (A0183156A), Naaman Tan (A0205299X), Chetwin Low (A0217618B)

Introduction: Image inpainting

Image inpainting refers to the repair of missing parts of an image, given one that is damaged or incomplete. Since the mapping between input and output images is inherently ill-posed given that many realistic and semantically possible solutions exist, image inpainting is a non-trivial problem and an active area of research [1]. Specifically, inpainting can be formulated as a conditional image generation problem, where a model takes as input a damaged image and outputs an image that is visually and semantically plausible image that is consistent with the input. The dominant approaches in image inpainting is to utilise Generative Adversarial Networks (GAN) or Variational Autoencoders (VAE) [2], or their variants.

In this project we aim to tackle the problem of image inpainting by building a model that generates images that are realistic and consistent with respect to the damaged input images.

Methodology

Approach

Mathematically, the image inpainting task can be expressed at the level of sets: we are given one set of images in domain X and a different set in domain Y, and can train a mapping f: $X \rightarrow Y$ such that the output images are indistinguishable from known images from the target distribution Y. We therefore seek to explore image-generation techniques that can learn to translate between domains, with the assumption that there is some underlying relationship between the domains. For example, that they are two different and reasonable renderings of the same underlying scene - and we explore techniques that can learn that relationship.

We can define the training task by the input, output and loss – our model takes as input an image that has random parts of it removed, reconstructs an image, and is penalised by some notion of reconstruction loss in terms of similarity to the ground truth image (e.g. L1 or L2 loss). The exact loss function will depend on model architecture (e.g. GANs' adversarial loss) and downstream decisions.

Dataset

iNaturalist [3] is an online social network that is also a crowdsourced species identification system and organism occurence recording tool. The total collection of verified images includes 91 million observations of 344 thousand species. We intend to scrape the website to build a balanced dataset of varying cats, dogs and bird species with a total of 20,000 images. We generate damaged images by utilising random crops and masks. In addition, we intend to construct multiple datasets with increasing difficulty levels to gradually verify the learning ability of our models. The properties to vary are: 1) image resolution, 2) types of animals in the dataset, and 3) number, shape, size, location of crops.

Evaluation metrics

Aside from qualitative evaluation, we will use metrics to measure image quality and diversity. Metrics for the former include Multiscale-SSIM [4], (peak) signal-to-noise-ratio [5], while those for the latter include Learned Perceptual Image Patch Similarity [6] to measure the diversity of generated images.

Project Milestones

Week	Milestone & Allocation
8	[Chetwin] Dataset scraping, pre-processing [Naaman] Dataset exploration [Yuan Bo] Build pipeline to generate multiple datasets with varying levels of difficulty
9	[Chetwin] Build MLP & CNN baselines and test on simplest dataset version
10	[Naaman] (Tentative) Improvement 1: Residual skips & dilated convolutions [Yuan Bo] (Tentative) Improvement 2: Image consistency with 2 discriminators [8]
11-12	[Together] (Tentative) Improvement 3: Explore attention [9], partial/gated convolutions [10]
13-14	[Together] Error analysis, ablation studies, and model interpretation [Together] Report writing and video presentation preparation

Z. Qin, Q. Zeng, Y. Zong, and F. Xu, “Image inpainting based on deep learning: A review,” Displays, vol. 69, p. 102028, 2021.

Y. Pang, J. Lin, T. Qin and Z. Chen, "Image-to-Image Translation: Methods and Applications," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2021.3109419.

“INaturalist,” *iNaturalist*. [Online]. Available: https://www.inaturalist.org/. [Accessed: 06-Mar-2022].

Z.Wang,E.Simoncelli, and A.Bovik,“Multiscale structural similarity for image quality assessment,”in *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003,* vol. 2, pp. 1398–1402 Vol.2, 2003.

A. Horé and D. Ziou, “Image quality metrics: Psnr vs. ssim,” in *2010 20th International Conference on Pattern Recognition*, pp. 2366–2369, 2010.

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *CVPR*, 2018.

P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2017.

S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Globally and Locally Consistent Image Completion,” ACM Transactions on Graphics (Proc. of SIGGRAPH), vol. 36, no. 4, pp. 107:1–107:14, 2017.

J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative image inpainting with contextual attention,” arXiv preprint arXiv:1801.07892, 2018.

J.Yu,Z.Lin,J.Yang,X.Shen,X.Lu,andT.S.Huang,“Free-form image inpainting with gated convolution,” arXiv preprint arXiv:1806.03589, 2018.