Image Quality Enhancement Using Machine Learning

Arushi Sinha, Jay Paranjape, Kwesi Buabeng Debrah, Sucheta Malladi

Problem Statement

- Increasing use of big data applications brings a need for downsizing the storage capacity that images take
- Domain generalization
  - Domain shift in source and target
    - Factors: hardware differences, light reflections, types of images
    - Challenge: Maintain performance across these domains
  - Using machine learning to extract key features from different types of images and being able to apply these across resizing operations across multiple different types of images
- Our target images: faces & endoscopic images

Goal

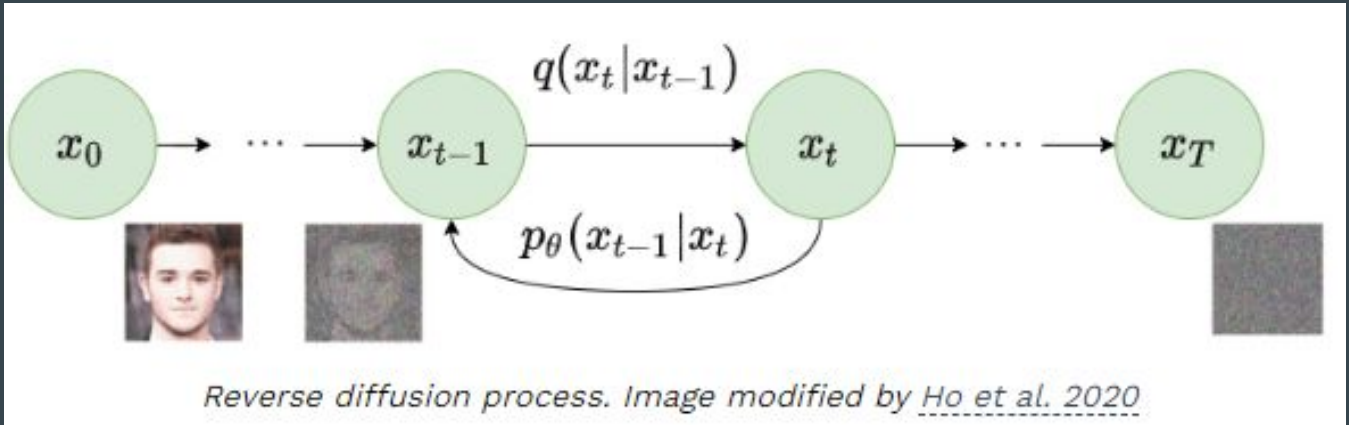
- Use machine learning to upscale face images and endoscopic images while maintaining image quality across multiple resizing operations
- Account for domain generalization
  - Train on face images and attempt to resize endoscopic images using the same algorithm
  - Train on endoscopic images and attempt to resize face images using the same algorithm
- Compare the results between the models and aim for low root mean squared error

Literature Review: Current Methods for Image Super Resolution

- CNN and attention based methods
  - Combines principles of transformers and U-Nets by modifying the attention structure of transformers
  - Makes the process of multi-headed self attention more efficient by attending over the channel dimension instead of the spatial dimension
- Using an enriched set of features that encompasses contextual and spatial information at various scales
  - Fine to coarse and coarse to fine semantic and spatial representations that are exchanges and attention based aggregated to extract attention based features
- SRCNN
  - Uses multiple convolution layers with ReLU activation
- Residual Dense Network (RDN)
  - Dense residual connections between SRCNN layers
- GAN: ESRGAN+
  - Adding noise to generator inputs to make the outputs stochastically realistic and adding dense residual networks (RDN)

An Overview of Diffusion Models

- Two parts - Encoder and Decoder. Encoder Learns the function q, Decoder Learns the function p
- Traditionally, at every step, gaussian noise is added. We also include blurring along with Gaussian noise
- Shared Weights, at every step the model gets image from previous timestep and t



Reverse diffusion process. Image modified by Ho et al. 2020

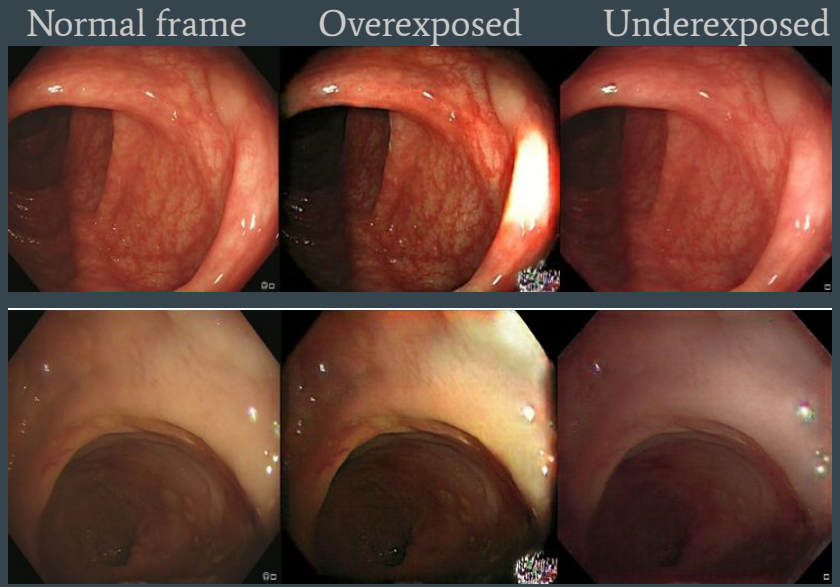
Proposed Method: Diffusion Model for Image Enhancement

- **Training:** train a diffusion model on dataset and use it to produce enhanced images
- **Hypothesis:** diffusion models will perform better on the same test set but might not produce good domain transfer results
  - Combining diffusion models for both training sets may learn better representations and perform better.
- **Example:** a diffusion model trained on faces would learn to generate faces from images with little added noise as well as images with a high amount of noise.
  - Generalize well on a test set from the same distribution.
  - Tradeoff: property might also make them weaker to domain shifts that could be mitigated using training data from multiple sources.

Dataset

- Human Faces Dataset
  - Decreased resolution to downscale images to half the scale and separated into training and testing using random 80:20 splitting

- Endo41E Dataset
  - Did not require additional processing or splitting



- Trained models on train set of one data set and then the test set of both data sets
  - Allows for measure of error on in distribution dataset and out of distribution dataset

Implementation

1. Baseline Models : RDN and SRCNN
2. Trained using the Endo41E training dataset and tested from both test datasets
3. Trained using the Human Faces training dataset and tested from both test datasets
4. Calculate RMSE between the original and reconstructed images
5. **Note:** Tried to implement the SOTA Restormer model

Initial Results and Analysis

RMSE	Model + Train Data			
Test Data	SRCNN on Faces	SRCNN on Endo41E	RDN on Faces	RDN on Endo41E
Faces	245971.72	199270.41	193022.70	223413.35
Endo41E (Over Exposed)	190230.43	162815.30	162904.87	171323.87
Endo41E (Under Exposed)	194169.04	183727.04	183175.88	191755.89

Analysis for Initial Results

- Overexposed Case: SRCNN trained on the Endo41E dataset produces the lowest error
- Overexposed case: RDN trained on faces produces the lowest scores
  - Counter intuitive result can be attributed to overfitting
- Increasing image resolution generalizes fairly well over these different datasets.
- Effect of domain transfer, while seen, is not as extreme as in other tasks

Results Obtained Using Diffusion Model

Test Data	Faces Model	Endo Model	Faces+Endo Model
Faces N20	172853.32	172813.48	172855.34
Faces N100	172853.36	172813.42	172855.44
Endo Under 20	168951.41	168952.34	168949.01
Endo Under 100	168951.28	168952.21	168949.21
Endo Over 20	165924.14	165695.45	165919.68
Endo Under 100	16592.14	165923.86	165919.63

Analysis - Diffusion Models for the Win!

- Intra Domain Error - Reduced Error when trained on faces/endo and tested on faces/endo
- Inter Domain Error - Reduced Error when trained on Faces/Endo and tested on Endo/Faces
- Speed - Fast!
    - Only requires <=20 timesteps to produce great images

What Does the Diffusion Model Learn During Training?

- Encoder Learns:

- Decoder Learns:

Image Generation for Endoscopy model

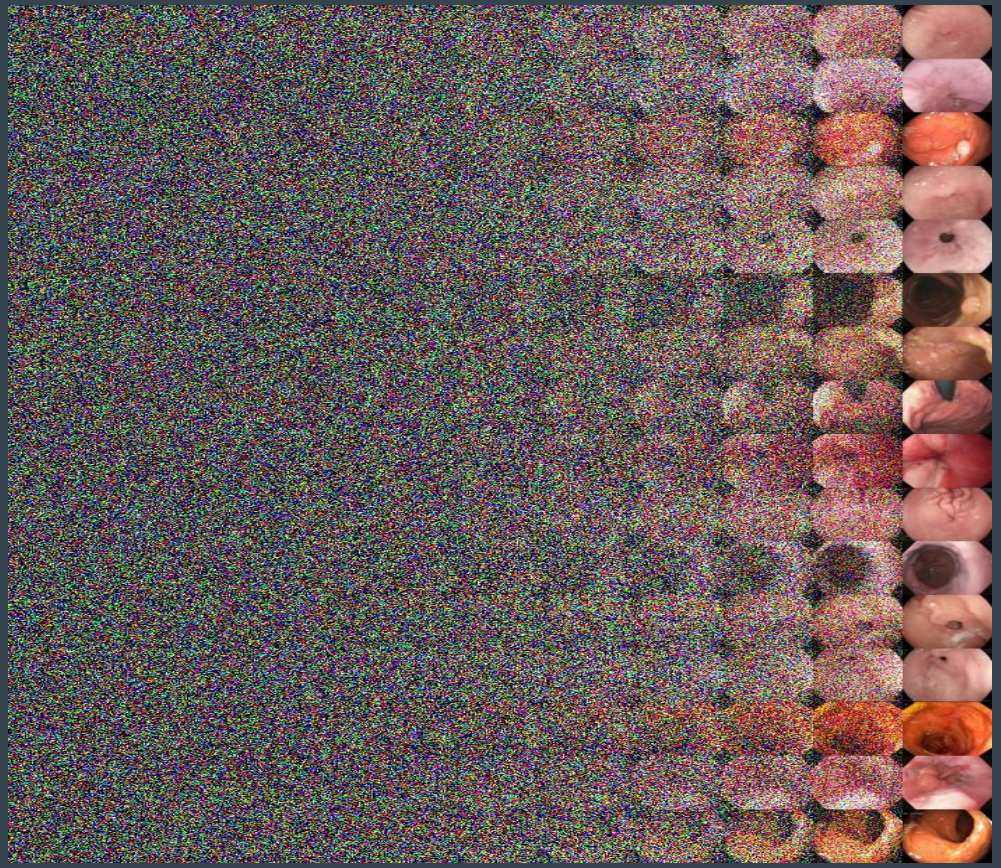
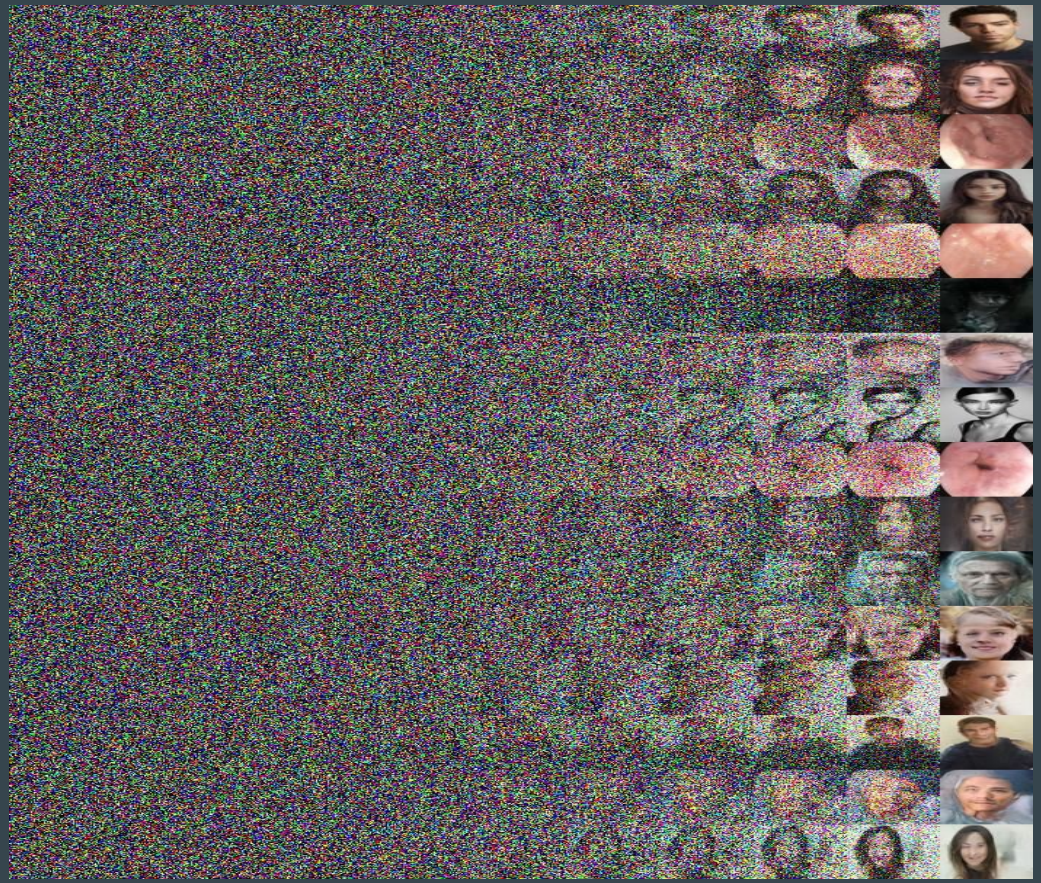


Image Generation for Combined (Faces + Endoscopy) Model

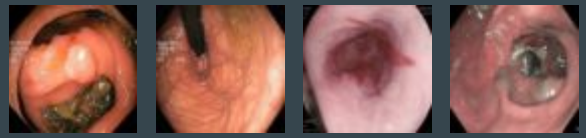


Visualization of the Testing Process

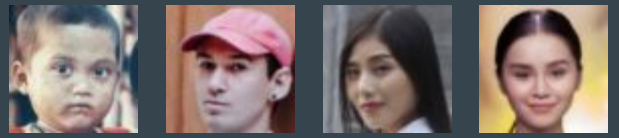
- Start with a non enhanced image - same as starting decoder at the very end.
- Hence, only requires a **few steps(N)** of the decoder
- We show that with as less as 20 steps, we get good results

Explaining Domain Generalization Capabilities

- Hypothesis - Diffusion Models trained on Faces would not work well on Endoscopy images
- Result:



Results(other way round)



- During testing, the decoder starts from the far right end. It takes as input the non enhanced image and the position.
- Hence, it learns how to deblur and denoise the input at that stage, which is agnostic of domain

Conclusion

- Improved Error within and Across Domains without affecting speed
- Explained the results with respect to our hypothesis
- Provided Visualizations for the Workings and Results of our method

References

[1] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang: “Restormer: Efficient Transformer for High-Resolution Image Restoration”, 2021; [http://arxiv.org/abs/2111.09881 arXiv:2111.09881]

[2] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Learning a Deep Convolutional Network for Image Super-Resolution, in Proceedings of European Conference on Computer Vision (ECCV), 2014 http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html

[3] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, Yun Fu: “Residual Dense Network for Image Super-Resolution”, 2018; [http://arxiv.org/abs/1802.08797 arXiv:1802.08797]

[4] Gupta, A. (2020). Human Faces, Version 1. Retrieved September 26, 2022 from https://www.kaggle.com/datasets/ashwingupta3012/human-faces

[5] Axel Garcia-Vega, Ricardo Espinosa, Gilberto Ochoa-Ruiz, Thomas Bazin, Luis Eduardo Falcon-Morales, Dominique Lamarque, Christian Daul: “A Novel Hybrid Endoscopic Dataset for Evaluating Machine Learning-based Photometric Image Enhancement Models”, 2022; [http://arxiv.org/abs/2207.02396 arXiv:2207.02396]

[6] S. W. Zamir et al., "Learning Enriched Features for Fast Image Restoration and Enhancement," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2022.3167175

[7] Ho, Jonathan and Jain, Ajay and Abbeel, Pieter, “Denoising Diffusion Probabilistic Models”, https://arxiv.org/abs/2006.11239

Thanks for Listening!

Questions?