Unrolled Optimization with Deep Priors Applied to Computer-Generated Images from 3D Interactive Environments

Patrick Xu[1], **Hao-Jen (Chris) Chien**[1], **Daniel Ahn**[1], **Ran Gong**[1], **and Hersh Joshi**[1]

[1]Student, UCLA

This manuscript was compiled on July 18, 2020

This report strives to expand upon the results of the paper, "Unrolled Optimization with Deep Priors" (1). First the unrolled optimization with deep priors framework is used to iterative solve for the latent image of a given, blurred input image for image restoration (IR). Results are shown to demonstrate the performance of unrolled optimization with deep priors (ODP) in (1) for devonvolving images subjected to Additive White Gaussian Noise (AWGN), motion blur kernels, and out-of-focus disk kernels. Results are shown to be within 3-4dB of prior work results using a small dataset of 200 images. Novel images are then generated by a 3D physics engine and photo-realistic rendering from the paper "VRKitchen: A 3D Dynamic Interactive Environment for General Computer Vision Research" (2). Uniform fog effects are then added to these images to observe the efficacy of the ODP network in image defogging scenarios compared to conventional deep learning based methods. The difficulty in characterizing fog effects as a Point Spread Function (PSF) kernel results in deep learning methods being much more effective for defogging than the ODP architecture.

Unrolled Networks | CNN | Imaging | Deconvolution | Optimization

1. Introduction

Inverse imaging problems involve reconstructing a latent image from images under a known physical image formation. These image restoration and inverse imaging problems are very useful in fields such as computational, imaging, computer vision, and medical imaging. Real-world images are frequently subjected to variety of imperfections- such as lens effects, noise, and blur- which result in the degradation of observed image quality. The goal of inverse imaging problems is to develop solutions and frameworks that allow for the resolution of these imperfections and artifacts. For images subjected to a known imperfection, problems can be framed as naive deconvolution problems, where a known kernel- such as a blur kernel- is convolved with a latent image to produce a blurred output. These problems are further complicated by the presence of noise, such as read noise or shot noise. These sources of noise increase the computational complexity of deconvolution and distort latent images produced by direct deconvolution.

There are two primary types of approaches to solving these deconvolution problems. The first approach attempts to solve for the latent image as the optima of an optimization problem. These methods are limited by their inability to incorporate complex learned models and statistics of natural images and thus struggle to solve complex imaging problems. Deep learning-based methods provide the framework to solve for these complex statistics, but lack a structured framework to incorporate prior knowledge into the image formation model. All knowledge that the model learns is solely learned during training.

Diamond et. al. (1) propose a framework that incorporates prior knowledge of the image into deep networks with comparable speed and noise levels to networks specialized for particular forms of noise and blur. This model provides an easy to train, high performance solution for a variety of inverse image problems. This framework is very robust compared to specialized networks and algorithms such as PBDW(3), PANO(4), FDLCP(5), ADMM(6), and BM3D-MRI(7), providing comparable or improved performance with a more generalizable architecture(1).

Recent development in computer graphics and virtual reality technologies have allowed for the use of photo-realistic rendering for indoor and outdoor environments using 3D physics engines and photo-realistic rendering. Using these engines with a VR systems allows for human participation in simulated environments. This has applications in i) allowing human teachers to perform demonstrations to train agents (i.e. learn from demonstration) and ii) allow users to collect multi-modal sensor data to train agents for other computer vision tasks. These simulated environments require training in environments with fog, hail, and other atmospheric aberrations. Unrolled optimization with deep priors can facilitate removing the effect of these atmospheric aberrations for training of the network.

We compare the performance of the ODP network with a convolutional neural network (CNN) architecture for defogging simulated images. We find that because of the random scattering and additive nature of fog effects, the ODP network cannot effectively incorporate prior knowledge of fog scattering into its model, and thus struggles to defog images. However, using only a CNN without incorporating prior knowledge resulted in high performance. We present initial results that demonstrate the effectiveness of CNN architectures with regards to defogging.

2. Background

MAP Estimation The framework employed in this report solves inverse problems through maximum-a-posterior (MAP) estimation using a Bayesian model. The model is based on the following conditional probabilities: $P(y|Ax)$ and $P(x;\theta)$ where x is the unknown, ground-truth image created from a prior distribution $\Omega(\theta)$ which is solved found through the ODP (optimization with deep priors) method used in this report. The variable A is the linear operator applied to the image by the image system and y is the actual image captured by the camera which is affected by a noise distribution $\omega(Ax)$

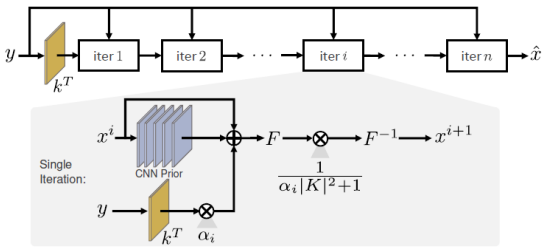

Fig. 1. Proximal Gradient ODP Algorithm

modeling sensor noise. $P(y|Ax)$ is the probability of sampling the image y from $\omega(Ax)$ and $P(x;\theta)$ is the probability of sampling the ground truth image from $\Omega(\theta)$

The MAP estimate of the ground-truth image x is given by following optimization equation where $f(y, Ax) = -\log P(y|Ax)$ and $r(x,\theta) = -\log P(x;\theta)$.

$$x = \arg\max_x f(y, Ax) + r(x, \theta), \quad [1]$$

Unrolled Optimization ODP attractively solves for the ground truth image x through a fixed, N, number of iterations. A non-constant number of iterations would allow for increased accuracy, but in practice this is not needed. This allows for the system to be viewed as an explicit function with inputs A, y, and θ outputting x^N. Each layer has a CNN and through chaining these in an unrolled algorithm this system can be viewed as a deep network.

Proximal Gradient The number of algorithm hyperparameters θ in the unrolled algorithm are usually small so the model capacity is based on the representation of the priorover the hyperparameters. The optimization method used here is not based directly on the prior r but instead seeks to minimize it through a proximal operator:

$$prox_{r(\cdot,\theta}(v) = \arg\min_z r(z,\theta) + \frac{1}{2}||z - v||_2^2 \quad [2]$$

Computer Vision in Simulated Environment Traditionally, visual representations are learned from static datasets. Either containing prerecorded videos (8) or images (9), most of them fail to capture the dynamics in viewpoint and object state during human activities, in spite of their large scale. Some early systems (10–13) try to simulate the dynamics of agent activities in order to support the development of smart visual surveillance systems and research into active computer vision for navigation. However, agents in the environment cannot be trained in a fined grained level for compositional tasks, and environments often do not have a lot of dynamic changes caused by agents' actions. However in real world scenarios where agents often need to navigation through different difficult environments to achieve his goals. This gives rise to the need of a simulated environment that is capable of simulating different extreme conditions.

Recent advances in the computer graphics and computer vision community create an opportunity to simulate difficult indoor and outdoor extremes. Sarker et al. (14) try to address this issue through the use of a system that is capable of simulate haze from a clear picture. However, such system does not capture agent's point of view which might serve as a input for artificial agents.

To address this issue, there has been a growing trend to develop 3D virtual platforms for training embodied agents in dynamic environments. Typical systems include 3D game environments (15–17), and robot control platforms (18–21). While these systems offer physics simulation and 3D rendering, they fail to provide realistic environments and daily tasks humans face in the real world. More recently, based on 3D scene datasets such as Matterport3D (22) and SUNCG (23), there are have been several systems simulating more realistic indoor environments (24–28) for visual navigation tasks and basic object interactions such as pushing and moving funitures (29). However, these system does not provide any capability to simulate extreme conditions that agents might face in daily lives.

Therefore, we believe to resolve this issue, a system that is capable of synthesising different extreme conditions (motion blur, smoke, fog, special lighting conditions) with multimodality data inputs in a dynamic environment is necessary. It naturally serve as a testbed for different algorithms.

3. Related Work

The ODP method within this papers is a re-implementation of the unrolled optimization with deep priors presented in the paper by Diamond et al.(1). The method is then extended to fog effects inspired by Gao's VR Kitchen paper(2).

Natural Image Priors Huang et al. (30) proposed that natural images often contains special statistics. Most dominant features are high kurtosis and sparsity. The image feature histogram is often Laplacian rather than Gaussian. Modeling natural image statistics has led different successful results in the past three decades. Zhu et al. (31) learn a Markov Random Field model to capture this image statistics in the texture synthesis domain. Yang et al. and Wang et al. (32, 33) use sparse coding for image classification.

Image Denoising through Image Priors Dong et al. (34) use sparse coding to learn a non-local sparse representation of the images and obtain promising results. Gu et al. (35) use weighted nuclear norm minimization to regularize images. However, these methods often suffer from the slow computation due to the optimization steps in the test-time. Recent works (36, 37) try to capture image priors through the use of deep neural network(mainly RNN and CNN).

Image Dehazing through Image Priors Similarly He et al. (38) use a simple haze-free image prior, haze-free images often have low intensity in at least one color channel, to perform haze removal task. Researchers show that haze images can be model as a linear combinations of direct attenuation and airlight contributions (39, 40). Based on these assumptions, Cai et al. (41) models the medium transmission priors through a CNN architecture. Yang et al. (42) propose to use a deep neural network to jointly model these priors for haze removal.

The success of these methods shows that modeling better assumptions and better priors does lead to a better performance.

Unrolled Optimization The unrolled optimization method may be expanded through parameterizing the prior to use learned sparsity priors. This entails representing the prior $r(x,\sigma)$ as $||Cx||_1$ where $Cx = (c_1 * x, ..., c_n * x)$ for convolutional kernels c composing a filterbank C (43, 44).

Field of Experts (FoE) Parameterizing the prior gradient as a field of experts through $g(Cx,\theta)$ g is a separable non-

linearity based on some θ such as a sum of radial basis functions and where C is once again a filterbank(6, 45–47). The ODP method outperforms the FoE method as FoE is in essence a 2-layer CNN as opposed to deeper CNN priors employed by ODP.

Deep models for Direct Image Inversion There are several methods employing CNNs developed to directly solve specific imaging issues. Xu et al. developed a network that deblurs within a single, learned deconvolution followed by a CNN(48). Schuler et al. propose a network that uses a single deconvolution step fed into a learned CNN(49). Wang et al. presents a CNN for MRI where the output is averaged with the raw values from the k-space (43).

4. Experiment Details

In this section we present results and analysis of ODP and CNN networks for denoising, deblurring, and defogging images.

A. Additive White Gaussian Noise. We consider the following problem for image formation with $y = x + z$, where the latent image x is subjected to Additive White Gaussian Noise (AWGN) z. The Bayesian estimation problem can be considered as the following optimatization problem

$$minimize \frac{1}{2\sigma^2}||x - y||^2 + r(x, \theta)$$

We trained a 5 iteration proximal gradient ODP network with an 10 layer, 64 channel residual CNN (RESNET) prior on 200 training images from the BSDS500 dataset provided by Martin et. al. (50). The images were edge tapered with a 10x10 Gaussian blur kernel to reduce edge artifacts during deconvolution and rotated 0°, 90°, and 180° to provide additional training data. Table 1 shows that our version of the ODP network provides comparable average test results to those presented by Diamond et. al (1).

Method	Gaussian Noise
BM3D(7)	28.56
EPLL (51)	28.68
Schmidt(46)	28.72
Wang (43)	28.79
Chen (47)	29.85
ODP (Diamond)(1)	29.04
ODP(Author's)	**29.5**

Table 1. Average PSNR (dB) for Images Corrupted by AWGN, $\sigma = 25$. CNN Method Uses No ODP or Proximal Gradient.

Visual assessment of the reconstructed image shows that the model can effectively suppress Gaussian noise with minimal impact on image quality. Due to the operations involved with denoising, high frequency details- such as object textures or dense object clusters- cannot be reconstructed perfectly. This is because noise is a high frequency phenomena, and noise suppression thus will result in the loss of high frequency details.

B. Deblurring. We consider the following problem of simultaneously deblurring and denoising. Here, the latent image x is convolved with a known Gaussian blur kernel and subjected to AWGN. The image formation model is $y = k * x + z$, where k is the blur kernel and z is AWGN. The Bayesian estimation problem can be expressed as

$$minimize \frac{1}{2\sigma^2}||k * x - y||^2 + r(x, \theta)$$

3 specific kernels were used to assess the performance of the ODP model: a Gaussian kernel with $\sigma = 1$, a motion blur kernel with $width = 21 pixels$, and a disk out-of-focus blur kernel with $radius = 7$.

We trained the a 5 iteration proximal gradient ODP network with 10 layer, 64 channel RESNET priors for all 3 kernels. The Gaussian blurred images were subjected to AWGN with $\sigma = 25$, and the motion and disk blurred images were subjected to AWGN with $\sigma = 5.7020$.

We also compare the performance of the ODP network to a 10 layer, 64 channel CNN model to assess the performance improvement offered by the unrolled proximal gradient step. These comparisons were conducted for motion and disk blur.

Gaussian Blur with Heavy AWGN We find that we can achieve 25.5dB PSNR for images blurred with a $\sigma = 1$ Gaussian blur kernel and subjected to $\sigma = 25$ AWGN. We can also achieve 27.5dB PSNR for images blurred with a $\sigma = 10.2$ Gaussian blur kernel and subjected to $\sigma = 7$ AWGN. This is comparable to the results presented by Diamond et. al. (1) for Gaussian blur kernels of $\sigma = 10.2$ and $\sigma = 2$. However, it is noted that these results were under AWGN of $\sigma = 5.7020$ and not $\sigma = 25$.

Method	Blur σ	AWGN σ	PSNR
ODP (Diamond)(1)	10.2	5.7020	24.76
ODP (Diamond)(1)	2	5.7020	27.23
ODP(Author's)	**1**	**25**	**25.5**
ODP(Author's)	**10.2**	**7**	**27.6**

Table 2. Average PSNR (dB) for Images Corrupted by Gaussian Blur

Visual assessment of the images shows that the majority of the image details have been maintained or recovered. However, the image lacks the sharpness present in the latent image and possesses a grainy texture. This is largely in part due to the large variance of the Gaussian noise, which has been largely suppressed or reduced in the reconstructed image. However, the reconstruction has resulted in the smoothing of noise, most noticeable against relatively monochromatic scenes. This results in the loss of high frequency details such as object textures.

Motion Blur with AWGN We find that we can achieve 24.9dB PSNR when reconstructing an image blurred by a 3x21 motion blur kernel. These results are 3.5dB below reported results for the original ODP model, but are comparable with the results presented by Krishnan (52), Levin (53), and Schuler (49).

We find that there is a 2.7dB increase in average PSNR for the set when deblurring with the ODP architecture compared to a CNN architecture. The CNN architecture provides minimal, if any, visual improvement or deblurring to the image. This validates the effectiveness of the proximal gradient algorithm and its ability to incorporate prior knowledge into a deep learning architecture.

Visual inspection of the reconstructed images shows that low frequency content, such as relatively constant objects and shapes, can be mostly recovered, while high frequency

Fig. 2. Example Latent Image (Left), Example Noised Image (Center), and Example Denoised Image (Right)

Fig. 3. Example Test Image (Left), Example Blurred and Noised Image (Center), and Example Deblurred and Denoised Image (Right), Gaussian Blur $\sigma = 10.2$, AWGN $\sigma = 7$

Fig. 4. Example Latent Image (Far Left), Example Motion Blurred Image (Left), Example CNN Deblurred Image (Right), and Example ODP Deblurred Image (Far Right)

content, such as sharp edges, cannot be entirely recovered. This is likely due to the unrolled algorithm's similarities to Weiner deconvolution, where high frequency content associated with noise is smoothed or suppressed to denoise an image. This results in the loss of edge details and sharpness in the reconstructed image.

Furthermore, there are notable edge artifacts, particularly around large objects and on the edges of the image. These edge artifacts can be partially attributed to the deconvolution process, which will produce ringing artifacts near image edges that are not tapered. To mitigate this to an extent, images were edge tapered with a 10x10 Gaussian kernel. The result is that the areas near the center of the image are nearly identical to the non-tapered image, and areas near the image edges resemble the Gaussian blurred images more.

Disk Blur with AWGN We find that we can achieve 25.4dB PSNR when reconstructing an image blurred by a 15x15 disk kernel. These results are comparable to the results presented by Diamond et. al. (1) as well as the results of related work, such as by Schmidt (46), and Krishnan (52).

We find there is a 3.6dB increase in average PSNR when using the ODP network compared to a standard CNN architecture. Similar to motion deblurring, a standard CNN struggles greatly when attempting to deconvolve blurred images without prior knowledge of the blur kernel.

Visual inspection of the reconstructed image shows that similar to motion deblurring, low frequency content is mostly recovered and high frequency content is not. Smoothing of sharp edges and dense object clusters is more severe for disk blur compared to motion blur. However, there are not as many

Fig. 5. Example Latent Image (Far Left), Example Disk Blurred Image (Left), Example CNN Deblurred Image (Right), and Example ODP Deblurred Image (Far Right)

significant motion or edge artifacts due to the isotropic nature of the disk blur kernel.

Method	Disk	Motion
Krishnan(52)	25.94	25.07
Levin(53)	24.54	24.47
Schuler(49)	24.67	25.27
Schmidt(46)	24.71	25.49
Xu(48)	26.01	27.92
ODP (Diamond)(1)	26.11	28.49
CNN	**22.3**	**22.2**
ODP(Author's)	**24.9**	**25.4**

Table 3. Average PSNR (dB) for Kernels with a 21x3 motion blur and disk blur $r = 7$, $\sigma = 5.7$. CNN Method Uses No ODP or Proximal Gradient.

C. Comparison of Model with Original ODP Model. There may be several potential sources of error resulting in decreased performance compared to the original presented ODP network. Our network was trained over 30000 iterations, while the network presented by Diamond et. al. (1) was trained over 130000 iterations. Furthermore, we fixed $\alpha_k = 0.5$, whereas Diamond. et. al. allow α_k to be a learnable parameter. In addition, our network does not reuse network parameters between iterations of the proximal gradient ODP algorithm. A new RESNET model is instantiated for each of the 5 iterations of the network. This provides some deviation from the original proposed model, where model parameters are reused over all iterations.

Furthermore, Gaussian denoising and motion deblurring experiments conducted by Diamond et. al.(1) were performed on grayscale images, whereas our experiments were performed on RGB images. This may result in different performance due to the presence of additional color channels.

In addition, deblurring experiments conducted by Diamond et. al. (1) were performed on 12000 training images from the ImageNet dataset. Images were also cropped to 256x256 images, rotated by 0°, 90°, and 180°, and subjected to AWGN with $\sigma = 5.7020$. We have mimicked the data rotation and dynamic noise generation, but have instead resized our latent images to 300x300 instead of taking random crops. The resizing of these images may have caused the image statistics to differ from natural images, as images by be compressed vertically or horizontally. In addition, the comparative lack of data in the BSDS500 dataset compared to the ImageNet dataset may result in less generalized network results and thus decreased performance.

Also, it was discovered that there may be a more optimal implementation of proximal gradient for RGB images. Currently, our implementation uses the algorithm outlined in Fig.1 across a 300x300x3 RGB image. Since many of the operations in this algorithm are performed in the frequency domain, we have found that we can achieve better results by performing the algorithm across each RGB channel separately. However, this has not been fully implemented due to time constraints.

D. Simulation Engine. In order to better simulate different extreme conditions as an input for our algorithms, we choose to use VRKitchen (2) as our base environment. VRKitchen is built using Unreal Engine 4(UE4). It provides exclusive utilities to simulate different conditions ranging from fog, smoke, motion blur etc. In this experiment, we choose two tasks (turn on light, cut meat) form VRKitchen and add fog on top of it to simulate our desired input as shown in Figure 6 and Figure 7. In addition to the example demonstrated, the environment is also capable of recording frames from a multi-view camera simultaneously at a frame rate about 20-24 FPS depending on the resolution of the images. Particle effects like smoke (Figure 8), snow etc are also easily synthesised.

E. Defogging. We initially consider the following problem of defogging a computer generated image similarly to the Bayesian estimating problem for deblurring. This initial assumption can be expressed as

$$minimize \frac{1}{2\sigma^2} ||k * x - y||^2 + r(x, \theta) \qquad [3]$$

This model assumes that the simulated fog effects can be characterized as a PSF kernel. The kernel was estimate generating an image matrix of a point source under fog effects in the simulated environment and is shown in Figure 9.

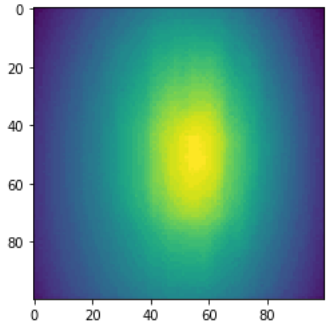

Fig. 9. Fog Kernel

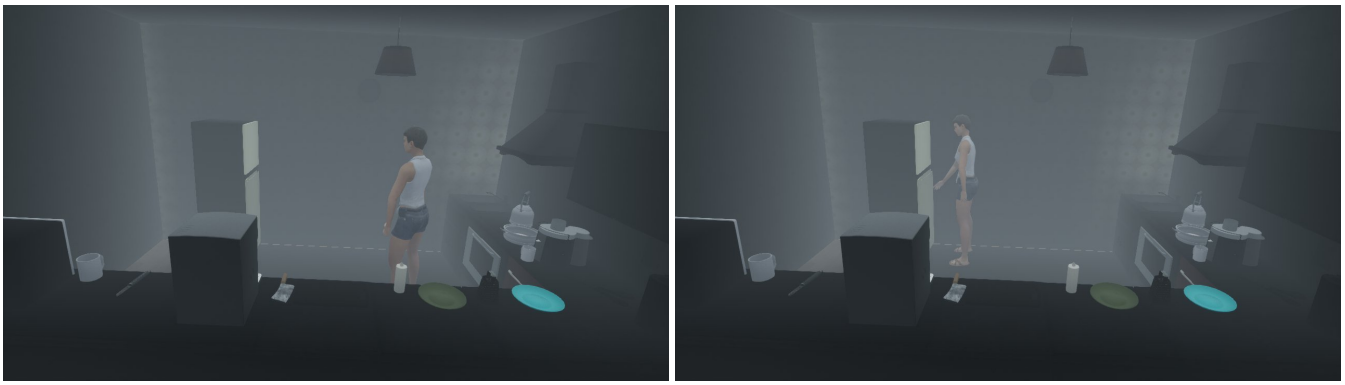

Fig. 6. Sample Synthesised Image Under Foggy Environment

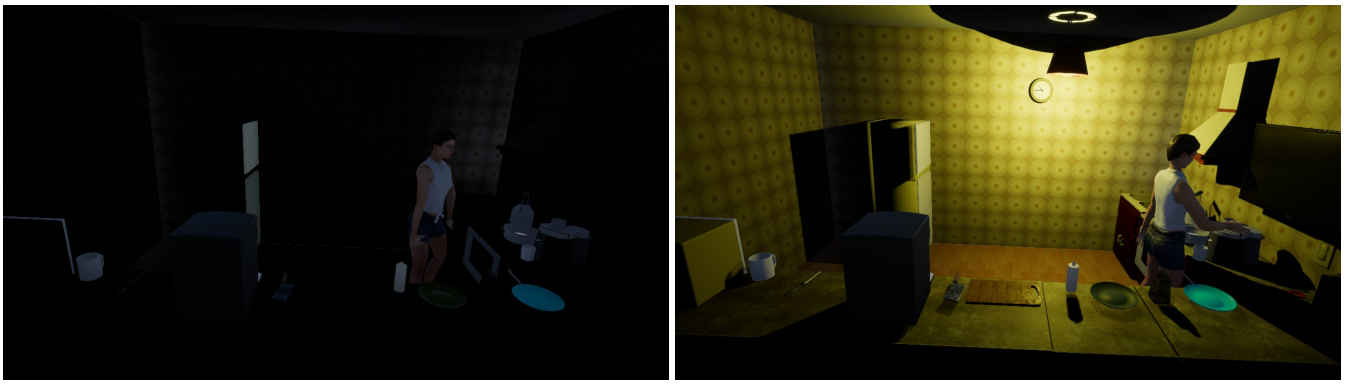

Fig. 7. Sample Synthesised Image Under Non-Foggy Environment

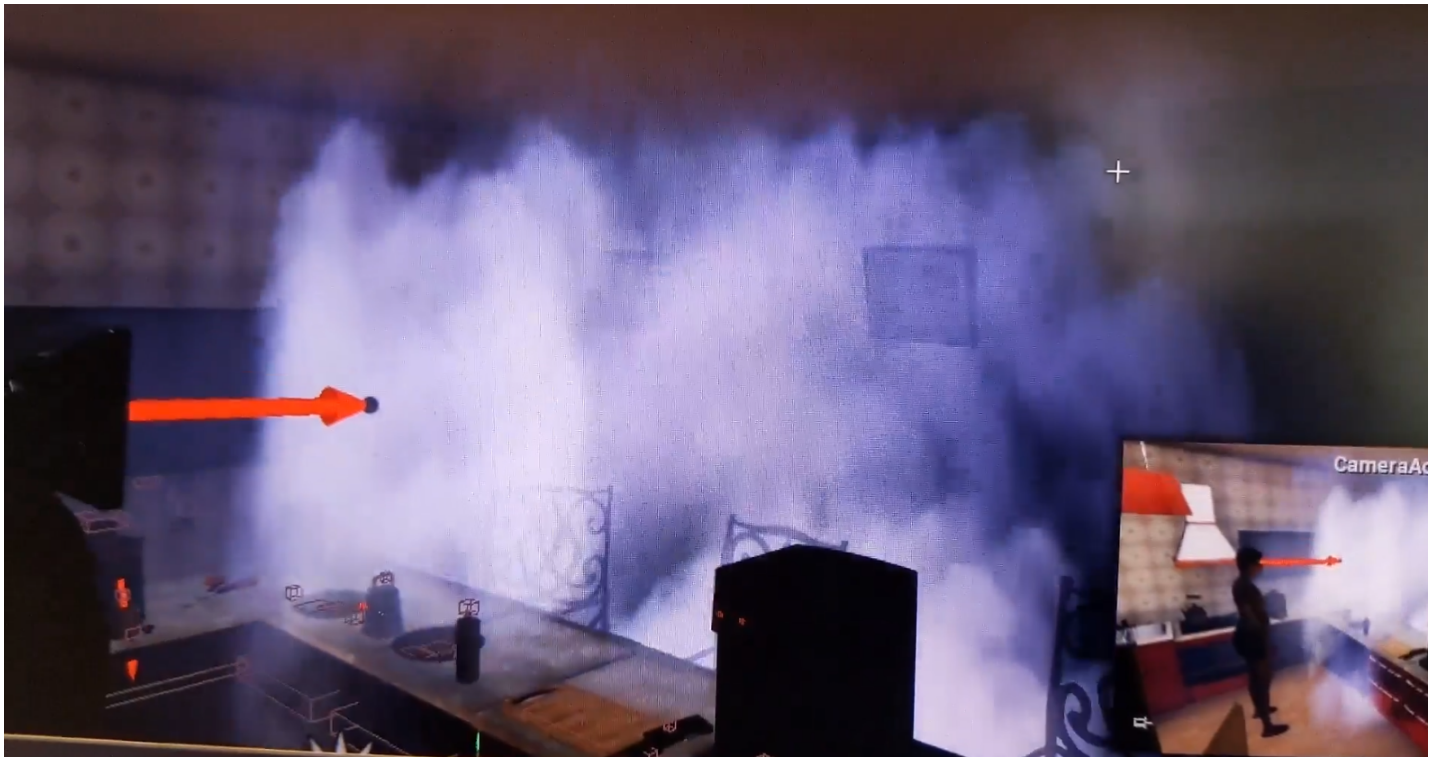

Fig. 8. Sample Synthesised Smoke Image

However, this estimated PSF has been found to be not representative of the full fog effects, this is because the heavy scattering by fog results in a very large kernel, on par with the size of the entire image. Due to computational limitations, our experiments were limited to testing a smaller fog PSF of size 10×10 show in Figure 10.

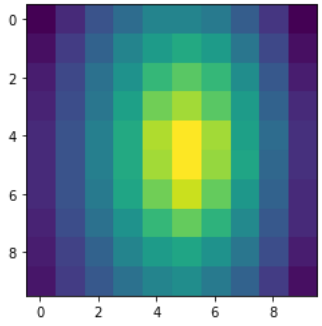

Fig. 10. Small Fog Kernel

Furthermore, the assumption that fog effects can be modelled as a fog kernel k convolved with a latent image x has been found to be erroneous. This is because fog adds additional information to an image, whereas a kernel only manipulates image already present in the image. Thus, it is not entirely accurate to assume that fog effects can be considered as a purely convolutional effect.

Method	Fog
Foggy Image	27.42
CNN	41.52
ODP(Author's)	-42.42

Table 4. Average PSNR (dB) for Fog Blur Kernel. CNN Method Uses No ODP or Proximal Gradient.

This assumption has been verified with the very poor performance of the ODP network with the estimated PSF. Because of the "washing-out" of the fogged image, the ODP network architecture over-saturates the reconstructed image, thus resulting in incredibly poor performance. The ODP network achieved -41dB PSNR when utilizing the estimated 10x10 PSF kernel. These results suggest that prior knowledge of weather effects cannot be easily integrated into existing image reconstruction models, such as ODP. The scattering and additive effects of fog cannot be easily estimated by the ODP architecture due to the inaccuracy of using a fog PSF as prior knowledge.

It was found however, that utilizing a conventional CNN to learn image statistics without help of a fog PSF provided greatly improved performance. We find that we can achieve 41.5dB average PSNR over a test set of 60 images, providing a 14dB average increase in PSNR. This massive performance improvement can likely be attributed to the PSF kernel. It is likely that the use of the fog kernel during the unrolled portion of the proximal gradient algorithm greatly distorts results and prevents the model from efficiently learning the image statistics of the latent images. However, if this kernel is removed, then a CNN is capable of learning and reconstructing defogged simulated images.

Using the fog data generated by the simulation engine, seen in Figure 11, we can see a comparison of the results of the ODP and CNN only method in Figure 12.

5. Conclusions

The proposed ODP framework presented by Diamond et al(1) has been replicated and validated for denoising and deblurring inverse imaging problems. The model provides an efficient and adaptable method of combining prior knowledge of the image formation model with deep learning models, allowing for improved performance compared to traditional optimization-based or CNN-based models. We have presented that we can achieve comparable results to that presented by Diamond et al(1) for motion and disk kernel deblurring.

Furthermore, we have analyzed the performance of the ODP and CNN-based models for a new image formation model involving defogging simulated images. We find that the difficulty in succinctly characterizing fog effects as a PSF kernel hinders the performance of the ODP model. However, we have also found that CNN-based models are capable of reconstructing images obscured by fog with very high reconstructed image quality. We have demonstrated the limitations of the ODP model in regards to reconstructing images subjected to effects other than traditional kernel-based image formation models and have presented results that suggest that for these purposes, deep learning models may be more efficient.

In the future, we would like to explore further optimizations that can be made to our ODP model to better match the results presented by the original ODP model. In particular, we are interested in further optimizing our implementation of proximal gradient to better suit RGB images, as the current implementation cannot attain the results achieved by Diamond et al(1). Furthemore, we would like to further analyze how to effectively incorporate complex weather effects such as fog into both CNN-based and ODP models to improve the generalizability and performance of these models.

6. Contributions

Patrick Xu Patrick was the primary member responsible for designing and building the CNN and ODP networks and several helper functions. He built, debugged, and evaluated the performance of the models for all denoising and deblurring experiments with the ODP and CNN models. He is also responsible for evaluating and debugging the performance of the models for defogging, and discovered that the CNN architecture is more effective than the ODP model for defogging. He authored the introduction, experimental results, and conclusion sections of this report.

Hao-Jen Chien Chris helped build helper functions and helped carefully debug the ODP model. She provided the edge tapered data sets in MATLAB, and also observed and helped debug different issues related to loading and different blurring/noising image data. She also proposed and provided data augmented image sets to help improve performance and helped evaluate the ODP model.

Daniel Ahn Daniel helped build and debug the ODP models and discovered that the CNN architecture is more effective than the ODP model for defogging. He built several helper functions, including functions related to loading and modifying test set data. He helped author sections on defogging and provided edits to the paper.

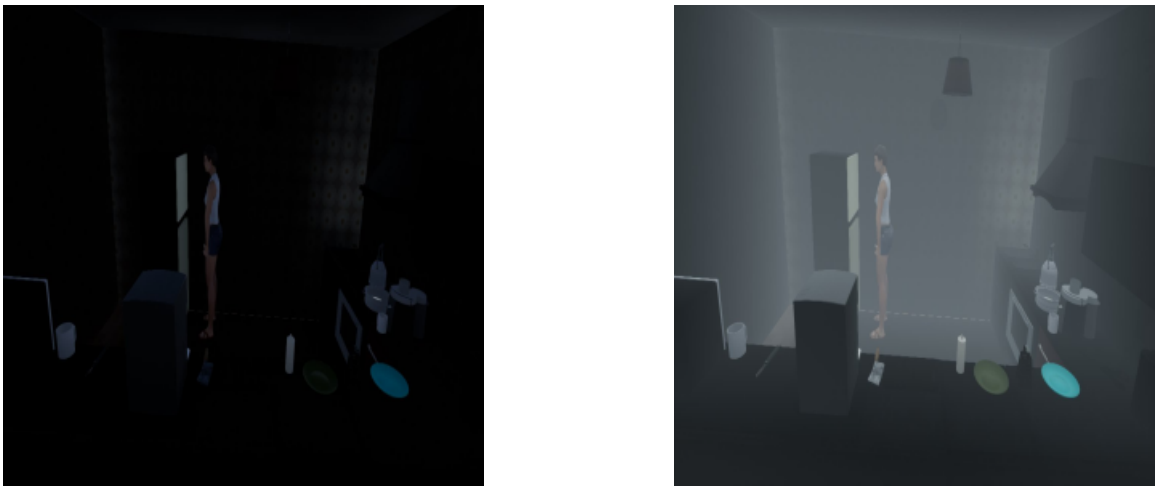

Fig. 11. Ground Truth Image (Left) and Fogged Image (Right)

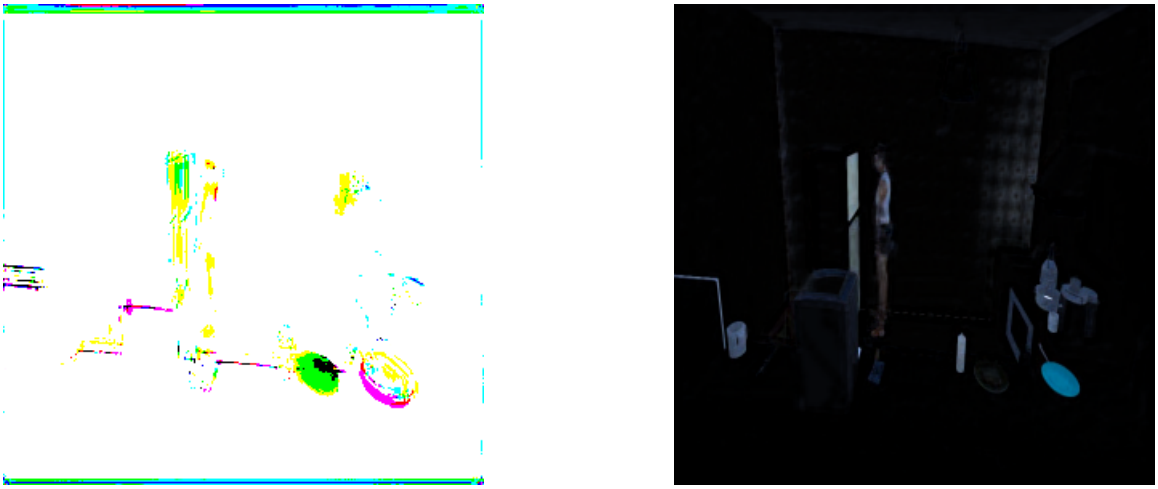

Fig. 12. Reconstructed Images with ODP (Left) and CNN (Right)

Ran Gong Ran proposed the use of the ODP model to defog computer generated images. He provided the fog PSF and computer generated test and training data to train the defogging models through the use of VRKitchen simulated environment. https://sites.google.com/view/vr-kitchen/. He coauthored the VRKitchen paper. He authored sections related to the simulation engine and background of the computer generated data of this report. He also did a literature review on the related work section

Hersh Joshi Hersh primarily worked on the report and helping break down the results of the initial paper. Hersh authored the abstract, background, and related work sections of the paper.

1. Diamond S, Sitzmann V, Heide F, Wetzstein G (2017) Unrolled optimization with deep priors.
2. Gao X, et al. (2019) Vrkitchen: an interactive 3d virtual environment for task-oriented learning. *arXiv preprint arXiv:1903.05757*.
3. Qu X, et al. (2012) Undersampled mri reconstruction with patch-based directional wavelets. *Magnetic Resonance Imaging* 30(7):964 – 977.
4. Qu X, et al. (2014) Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator. *Medical image analysis* 18(6):843–856.
5. Zhan Z, et al. (2015) Fast multiclass dictionaries learning with geometrical directions in mri reconstruction. *IEEE Transactions on Biomedical Engineering* 63(9):1850–1861.
6. Sun J, Li H, Xu Z, , et al. (2016) Deep admm-net for compressive sensing mri in *Advances in neural information processing systems*. pp. 10–18.
7. Danielyan A, Katkovnik V, Egiazarian K (2011) Bm3d frames and variational image deblurring. *IEEE Transactions on Image Processing* 21(4):1715–1728.
8. Rohrbach M, Amin S, Andriluka M, Schiele B (2012) A database for fine grained activity detection of cooking activities in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 1194–1201.
9. Jia Deng, et al. (2009) ImageNet: A large-scale hierarchical image database in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 248–255.
10. Qureshi F, Terzopoulos D (2008) Smart camera networks in virtual reality. *Proceedings of the IEEE* 96(10):1640–1656.
11. Rabie TF, Terzopoulos D (2000) Active perception in virtual humans in *Vision Interface*. Vol. 2000.
12. Terzopoulos D, Rabie TF (1995) Animat vision: Active vision in artificial animals in *Proceedings of IEEE International Conference on Computer Vision*. (IEEE), pp. 801–808.
13. Lin J, et al. (2016) A virtual reality platform for dynamic human-scene interaction in *SIGGRAPH ASIA 2016 virtual reality meets physical reality: Modelling and simulating virtual humans and environments*. (ACM), p. 11.
14. Sarker A, Akter M, Uddin MS (2019) Simulation of hazy image and validation of haze removal technique. *Journal of Computer and Communications* 7:62–72.
15. Kempka M, Wydmuch M, Runc G, Toczek J, Jaskowski W (2017) ViZDoom: A Doom-based AI research platform for visual reinforcement learning. *IEEE Conference on Computatonal Intelligence and Games, CIG*.
16. Beattie C, et al. (2016) DeepMind Lab. pp. 1–11.
17. Johnson M, Hofmann K, Hutton T, Bignell D (2016) The malmo platform for artificial intelligence experimentation. *IJCAI International Joint Conference on Artificial Intelligence* 2016-Janua:4246–4247.
18. Todorov E, Erez T, Tassa Y (2012) MuJoCo: A physics engine for model-based control in *IEEE International Conference on Intelligent Robots and Systems*.
19. Coumans E, Bai Y (2016) Pybullet, a python module for physics simulation for games, robotics and machine learning. *GitHub repository*.
20. Fan L, et al. (2018) SURREAL: Open-Source Reinforcement Learning Framework and Robot

Manipulation Benchmark.
21. Plappert M, et al. (2018) Multi-Goal Reinforcement Learning: Challenging Robotics Environments and Request for Research.
22. Chang A, et al. (2018) Matterport3D: Learning from RGB-D data in indoor environments. *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017* pp. 667–676.
23. Song S, et al. (2017) Semantic scene completion from a single depth image in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Vol. 2017-Janua, pp. 190–198.
24. Brodeur S, et al. (2017) HoME: a Household Multimodal Environment. No. Nips, pp. 1–6.
25. Wu Y, Wu Y, Gkioxari G, Tian Y (2018) Building Generalizable Agents with a Realistic and Rich 3D Environment.
26. Savva M, Chang AX, Dosovitskiy A, Funkhouser T, Koltun V (2017) MINOS: Multimodal Indoor Simulator for Navigation in Complex Environments. pp. 1–14.
27. McCormac J, Handa A, Leutenegger S, Davison AJ (2017) SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-training on Indoor Segmentation? in *Proceedings of the IEEE International Conference on Computer Vision*. Vol. 2017-Octob, pp. 2697–2706.
28. Xia F, et al. (2018) Gibson Env: Real-World Perception for Embodied Agents in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
29. Kolve E, et al. (2017) AI2-THOR: An Interactive 3D Environment for Visual AI. pp. 3–6.
30. Huang J, Mumford D (1999) Statistics of natural images and models in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*. (IEEE), Vol. 1, pp. 541–547.
31. Zhu SC, Wu Y, Mumford D (1998) Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *International Journal of Computer Vision* 27(2):107–126.
32. Yang J, Yu K, Gong Y, Huang T (2009) Linear spatial pyramid matching using sparse coding for image classification in *2009 IEEE Conference on computer vision and pattern recognition*. (IEEE), pp. 1794–1801.
33. Wang J, et al. (2010) Locality-constrained linear coding for image classification in *2010 IEEE computer society conference on computer vision and pattern recognition*. (Citeseer), pp. 3360–3367.
34. Dong W, Zhang L, Shi G, Li X (2012) Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing* 22(4):1620–1630.
35. Gu S, Zhang L, Zuo W, Feng X (2014) Weighted nuclear norm minimization with application to image denoising in *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2862–2869.
36. Liu D, Wen B, Fan Y, Loy CC, Huang TS (2018) Non-local recurrent network for image restoration in *Advances in Neural Information Processing Systems*. pp. 1673–1682.
37. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* 26(7):3142–3155.
38. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence* 33(12):2341–2353.
39. Fattal R (2008) Single image dehazing. *ACM transactions on graphics (TOG)* 27(3):72.
40. Tan RT (2008) Visibility in bad weather from a single image in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. (IEEE), pp. 1–8.
41. Cai B, Xu X, Jia K, Qing C, Tao D (2016) Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing* 25(11):5187–5198.
42. Yang D, Sun J (2018) Proximal dehaze-net: a prior learning-based deep network for single image dehazing in *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 702–717.
43. Wang S, Fidler S, Urtasun R (2016) Proximal deep structured models in *Advances in Neural Information Processing Systems*. pp. 865–873.
44. Gregor K, LeCun Y (2010) Learning fast approximations of sparse coding in *Proceedings of the 27th International Conference on International Conference on Machine Learning*. (Omnipress), pp. 399–406.
45. Roth S, Black MJ (2005) Fields of experts: A framework for learning image priors in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. (Citeseer), Vol. 2, pp. 860–867.
46. Schmidt U, Roth S (2014) Shrinkage fields for effective image restoration in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2774–2781.
47. Chen Y, Yu W, Pock T (2015) On learning optimized reaction diffusion processes for effective image restoration in *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 5261–5269.
48. Xu L, Ren JS, Liu C, Jia J (2014) Deep convolutional neural network for image deconvolution in *Advances in neural information processing systems*. pp. 1790–1798.
49. Schuler CJ, Christopher Burger H, Harmeling S, Scholkopf B (2013) A machine learning approach for non-blind image deconvolution in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1067–1074.
50. Martin D, Fowlkes C, Tal D, Malik J, , et al. (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. (Iccv Vancouver:).
51. Zoran D, Weiss Y (2011) From learning models of natural image patches to whole image restoration in *2011 International Conference on Computer Vision*. (IEEE), pp. 479–486.
52. Krishnan D, Fergus R (2009) Fast image deconvolution using hyper-laplacian priors in *Advances in neural information processing systems*. pp. 1033–1041.
53. Levin A, Fergus R, Durand F, Freeman WT (2007) Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)* 26(3):70.