

國立臺灣大學電機資訊學院資訊網路與多媒體研究所



碩士論文

Graduate Institute of Networking and Multimedia
College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

利用多目標生成對抗網路去除單張影像反射現象

Single Image Reflection Removal based on a Multi-Task

Generative Adversarial Network

周念新

Nien-Hsin Chou

指導教授：李明穗 博士

Advisor: Ming-Sui Lee, Ph.D.

中華民國 106 年 7 月

July 2017

國立臺灣大學碩士學位論文
口試委員會審定書

利用多目標生成對抗網路去除單張影像反射現象
Single Image Reflection Removal based on a Multi-Task
Generative Adversarial Network

本論文係周念新君（學號 R03944029）在國立臺灣大學資訊網路與多媒體研究所完成之碩士學位論文，於民國一百零六年七月廿五日承下列考試委員審查通過及口試及格，特此證明

口試委員：

李明穗

（簽名）

（指導教授）

李昇彥

周承德

楊佳玲

所長：

中文摘要



去除影像中的反射現象一直是影像處理與電腦視覺中公認的難題之一。然而其所帶來的影響，輕則讓留有瞬間珍貴回憶的相片留下瑕疵，嚴重甚至會影響電腦視覺系統的運作，如監視系統等。因此一個可於合理時間與空間成本下運行之有效去反射的方法成為了許多研究者想要追求的目標。

在這本篇論文中，我們提出一個基於多目標生成對抗網路的學習架構來學習如何在合理的時間內盡可能地去除影像中的反射影像。透過修改原始生成對抗網路之損失函數與加入多目標學習架構，並配合資料生成模型產生出足夠支持學習架構之資料集使網路能學習到具母體代表性之參數組合。

相較於現有以最佳化為基礎的去反射演算法，我們的方法有兩大優勢：非常快的執行速度與超越現有方法之去反射結果。透過生成對抗網路的特性，我們成功地在去除反射的同時，保留了其他方法無法留下的材質與細節，大大提高了結果影像的品質。

關鍵字：影像去反射、生成對抗網路、多目標學習

ABSTRACT



Removing reflection from the image is one of the hardest problem in the computer vision and image processing. However, its impact to our precious photos or computer vision system, like surveillance camera, is significant and desired to be solved. Therefore, we aimed to find out a practical approach to accomplish the removal in reasonable time and space.

In this thesis, we proposed a new network architecture called Multi-Task Generative Adversarial Network. We trained network to learn how to remove reflection as much as possible and in proper time as well. In order to support learning, we proposed a data synthesis model that synthesize realistic reflection-containing images. By modifying loss function and imposing multi-task architecture into network, we expected our network can learning the parameters to separate reflection layer from background.

Compared to major optimization-based algorithms, our network had two advantages: fast and effective. We successfully removed reflection and kept the texture and detail as well. By taking the advantage of Generative Adversarial Network, the quality of result has been significantly improved.

Key words: Image Reflection Removal, Generative Adversarial Network, Multi-Task Learning.

CONTENTS



口試委員會審定書	#
中文摘要	i
ABSTRACT	ii
CONTENTS	iii
LIST OF FIGURES	iv
Chapter 1 Introduction.....	1
Chapter 2 Related Work.....	4
2.1 Multiple Images Reflection Removal.....	4
2.2 Single Image Reflection Removal	5
Chapter 3 Multi-Task Generative Adversarial Network on Single Image Reflection Removal	7
3.1 System Overview.....	7
3.2 Data Synthesis Model	10
3.3 Multi-Task Generative Adversarial Network.....	14
Chapter 4 Experimental Results.....	18
4.1 Experiment Settings.....	18
4.2 Qualitative Results.....	18
4.3 Compared with other single-image-based approaches	20
Chapter 5 Conclusion and Future Work.....	23
5.1 Conclusions	23
5.2 Future Work	23
REFERENCE	25

LIST OF FIGURES



Fig. 3.1 A photo with reflection.....	9
Fig. 3.2 Comparison between additive model and screen model	10
Fig. 3.3 System flowchart.....	10
Fig. 3.4 Illustration of energy decay on reflection light	12
Fig. 3.5 Flowchart of data synthesis model	12
Fig. 3.6 Surface regression of intensity in different energy ratio	13
Fig. 3.7 Examples of synthetic data.....	14
Fig. 3.8 Architecture of image-to-image Generative Adversarial Network [15].....	16
Fig. 3.9 Architecture of Multi-Task Generative Adversarial Network.....	17
Fig. 3.10 Intermediate output and composited result.	17
Fig. 4.1 Quality results of successful case.....	19
Fig. 4.2 Quality results of some interesting case.....	20
Fig. 4.3 Comparison of other single-image-based approaches.....	21
Fig. 4.4 Results of iterative removal on strong reflection	22


Chapter 1 Introduction



There are various problems we would met when observing this world through the camera. For example, overexposure or underexposure, defocusing, motion blurring or occlusion. All these problems not only make our precious photo become defective but also affect the operating of computer vision system. Hence, the researchers put a lot of effort into these problems constantly. Image reflection removal, as one of these tough problems, still do not have an effective way to solve.


When taking a photo on the target through a transparent medium like glass, water or acrylic sheet, the annoying reflection always appeared in the photo. When The light on the same side with camera hit the medium surface, partial energy pass through the surface called refraction, the remain bounce back called reflection. The ratio of refraction and reflection is decided by the medium property. The light composited of reflected light and transmitted light from the target through the medium would finally emit into camera lens and the scene on the same side with camera is captured in the image. The scene in the image has to be removed because it would be the interference of the coming image processing. This process is so-called image reflection removal.

So far, the approach on reflection removal can be categorized by the input amount: multiple images and single image. Multiple image approach need more than one image



as input. These input images are restricted to specific condition, such as movement direction [6, 7, 8, 9, 10], flash or non-flash [13], etc. The approaches worked on these conditional input images have retrieved significant results. For example, Xue et al. [8] impose parallax clue, dense motion fields, to separate reflection from background. It not only free the motion direction limit but also estimate reflection layer clear enough to recognize detail. However, the assumptions on input images in these approaches make taking the photo more difficult, e.g., the target object must be static. Therefore, these outstanding approaches are unlikely to put into practice. Comparing to the restriction of multiple images approaches, single image approaches have no rigid assumption and only need single image as input. This kind of approach is more practical but still has long way to go. Single image approaches are so far not effective due to the limited information from single input. Arvanitopoulos et al. [5], state-of-the-art in single image, use reflection suppression as topic instead of reflection removal highlighting the difficulty in single image.

Recent years, deep learning has been active research topic thanks to the progress of computing power. Many critical issues of image processing and computer vision have been achieved great success by the deep learning. Generative adversarial network, proposed by Goodfellow et al. in 2014 [14], has been approved it's significant performance in generating nature image. Isola et al. [15] leverage Generative



Adversarial Network to handle image-to-image translating problem. By providing data pair of desiring translation, Generative Adversarial Network can learn to perform similar translation. It can be apply to many image-to-image problem due to its flexibility on architecture. We proposed an improved version of image-to-image Generative Adversarial Network to solve reflection removal problem. In addition, we proposed a data synthesis system to generate sufficient amount of data to support learning process. Result shows that our quality is better than state-of-the-art approach on single image.

Chapter 2 Related Work



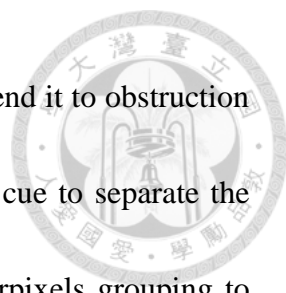
Previous research in reflection removal will be briefly introduce into two parts.

First, we present the recent approaches that require video or over two images as input.

Second, we list the single image approaches that is the category of our thesis.

2.1 Multiple Images Reflection Removal

Benefit from the rich information multi-images provided, researches on multiple images reflection removal never stop its step. Many researchers leverage a variety of clues to separate reflection from background. Most of methods focus on how to find the layer correlation between different images. Li and Brown [6] assume that multiple photos are taken in different angle so that reflection will vary in each images. Li et al. tried to relief user loading from [1] by using SIFT-flow to align transmission layer. After transmission layer is aligned, the edge on reflection layer can be classified. Then reflection layer can be estimated by the method in [1]. Guo et al. [7] used three priors to find optimized separation: correlation of transmission layer in images, the sparsity of the gradient fields of the two layers and the independence between the gradient fields of the transmitted and the reflected layers. Xue et al. [8] use motion difference to find the optimized dense motion field of two layers and separate the layer based on them




iteratively. Xue et al. free the constraint on motion direction and extend it to obstruction removal problem. Sun et al. [9] combine intensity cue and motion cue to separate the edge from input images. By using multi-level SIFT-flow and superpixels grouping to improve edge labeling, the amount of images required is relaxed to only two images. Yang et al. [10] proposed a new two-frame algorithm for robustly recovering the flow fields without restrictive assumption due to brightness constancy constraint.

Lai et al. [11] try to achieve goal in very different thinking. Unlike time-consuming optimizing approaches, they utilized fixed rank RPCA to separate low resolution images and then coupled dictionary for reconstruction of high resolution image.

2.2 Single Image Reflection Removal

Contrast to multiple images reflection removal, single image approaches have been bogged down for years. Levin and Weiss [1] proposed a classical method to solve single image layer separation by utilizing Laplacian mixture prior over image gradients. However, their method required user interaction to help labeling reference edge of two layers. Li and Brown [2] try to use smooth gradient prior on reflection layer and sparse gradient prior on transmission layer to replace user interaction in [1]. They make an assumption that the reflection layer always defocused and blurring than transmission layer, which is hardly valid due to the improvement of camera. Shih et al. [3] leveraged



the ghost effect on the reflection to achieve separation. They observed and found the reflection has two identical image overlapping on the transmission layer due to the thickness or plurality of glass. According to the ghost phenomenon, they use deblurring-based optimization to separate reflection from transmission. Wan et al. [4] believed transmission is in the Depth of Field (DoF) and reflection is out of the DoF. Therefore, they constructed DoF confidence map by Kullback-Leibler divergence to separate edge in transmission layer from reflection. Arvanitopoulos et al. [5] proposed a reflection suppression algorithm based on a l_0 gradient sparsity prior and a Laplacian difference. However, reflection and sharpness of transmission become the tradeoff in this work. Nevertheless, it is still the state-of-the-art in single image reflection removal and our competitor in this thesis.

Chapter 3 Multi-Task Generative Adversarial Network on Single Image Reflection Removal



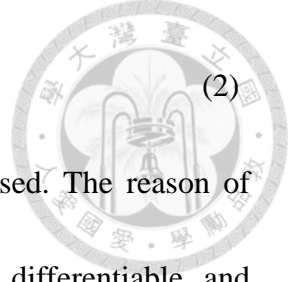
There are three parts in this chapter. Section 3.1 is about System overview, architecture and key model. Section 3.2 give a detail description of data synthesis model. Section 3.3 shows architecture about Multi-Task Generative Adversarial Network and related property.

3.1 System Overview

Single image reflection removal is an ill-posed problem. Information can be retrieved from single image is too insufficient to estimate accurate result. We assume reflection layer and transmission layer are composited by specific function. Formula (1) represent the assumption.

$$I = f(T, R) \quad (1)$$

$I \in \mathbb{R}^{n \times m}$ is the observed image, $T \in \mathbb{R}^{n \times m}$ is transmission layer, $R \in \mathbb{R}^{n \times m}$ is reflection layer and f represents the compositing function. So far, additive model is widely used in most of the approaches either single image or multiple images. The formula is as follows:



$$I = f(T, R) = T + R \quad (2)$$

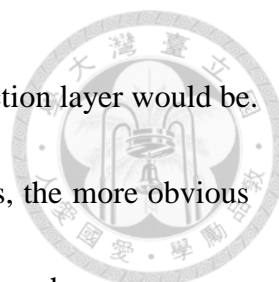
For given I , solving equation to estimate T and R still be ill-posed. The reason of widely used of additive model is its two excellent property: differentiable and replaceable. The unknown T can be easily replace by $I - R$ to reduce the complexity. These two property play an important role in optimization, which is the main method to solve reflection removal problem in most of the approaches. However, [12] mentioned that the compositing of reflection layer and transmission layer is decided by many factors instead of pure additive model. It is foreseeable and reasonable the result under additive model assumption would be unsatisfied. Hence, the present approaches always work like a charm on synthetic data but underperform on real image.

Instead of additive model, we assume that the compositing function is screen model. Screen model is, as additive model, one of blend model that is common build-in in most of the image processing software. The difference between screen model and additive model is that the layers are no longer independent. The formula of screen model is as follow:

$$I = f(T, R) = 1 - (1 - T) \odot (1 - R) \quad (3)$$

$$= T + (1 - T) \odot R \quad (4)$$

where \odot denotes element-wise multiply. In (4), reflection layer would multiply the weight, which is the invert of transmission layer before compositing. It means that the



stronger the region on the transmission layer is, the weaker the reflection layer would be. On the other hand, the darker the region on the transmission layer is, the more obvious the reflection layer become. The characteristic is in accord with the phenomenon we observed in real image. In fig. 3.1, the red region is sky, which has higher intensity and barely can see the reflection. On the contrary, the blue region is relatively dark and therefore reflection is clear.

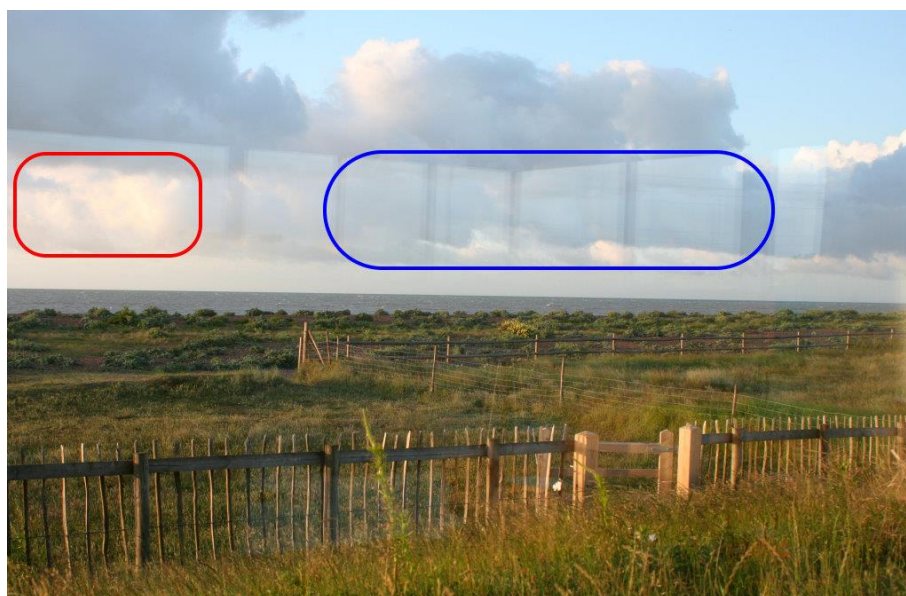


Fig. 3.1 A photo with reflection. Best viewed on screen.

In addition, we make an experiment to compare two model. In fig. 3.2, fig. 3.2a is an image with reflection and fig. 3.2b is transmission layer image taken by removing the glass. Fig. 3.2c is reflection layer image taken by putting pure black curtain on the back of the glass. All three images mention above is real images. Fig. 3.2d is the compositing result of fig. 3.2b and fig. 3.2c by additive model. Fig. 3.2e is compositing



by screen model. It is obvious to say that the result of screen model is more representative than additive model to the real image.



Fig. 3.2 Comparison between additive model and screen model. Best viewed on screen.

Based on screen blend model, the whole system is separated into two parts. System flowchart is as fig. 3.3. The first part is data synthesis model described in section 3.2 and the second part is learning model, described in section 3.3.

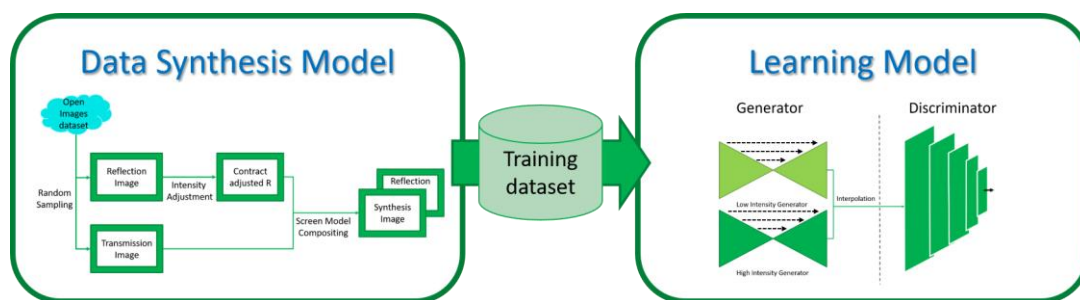
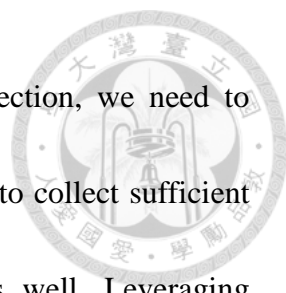


Fig. 3.3 System flowchart.

3.2 Data Synthesis Model

In the factors of effectiveness in machine learning, the amount of data should be at



the top. In order to make network learned the separation of reflection, we need to provide data as many as possible. However, it is nearly impossible to collect sufficient data pair in the real world and there is no existing dataset as well. Leveraging learning-based method to solve reflection removal problem is impractical.

Due to the difficulty of collecting data from real world, we decided to use synthesis way instead. Nevertheless, it is meaningful only when the synthetic data is similar enough to the real data. We exam the theory of reflection and related material to induce two major characteristics of reflection layer. First is the phenomenon of reflection layer on different intensity of transmission layer. This is described in section 3.1. Second, the energy decay in reflection layer. [3] mention that the intensity of reflection layer is the sum of the lights reflected from the medium surface. In fig. 3.4, according to Fresnel Equation, the energy of reflection is the R times of the energy of incident light whenever it reflected. R is so-called reflectance of the given medium. The reflectance of glass is usually between 0.05 and 0.1. Assume the energy of incident light T is E and reflectance of given medium is 0.1. The energy of first time reflection R_1 is $0.1E$ and the second time R_2 is $0.081E$. The energy after third time reflection R_3 ($<0.001E$) is negligible to the total energy of reflection. The total energy of reflection is around $0.181E$. It means that only 18.1% energy of incident light is reflected and it is around 9.5% left when reflectance is 0.05. Therefore, we need to decay intensity of reflection

layer to 9% to 18% when synthesizing data.

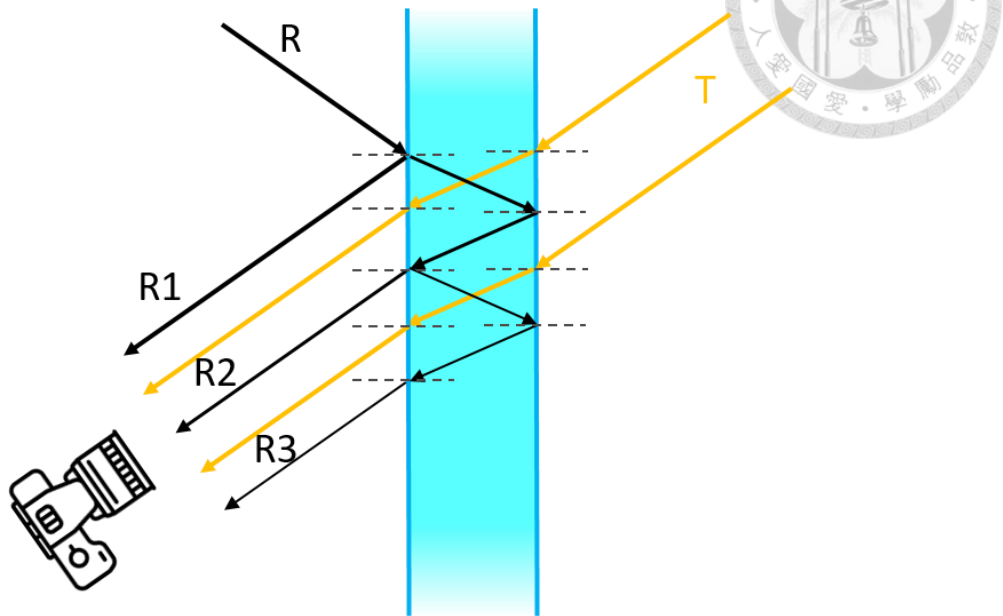


Fig. 3.4 Illustration of energy decay on reflection light.

We assume the origin mean intensity of reflection layer is equal or less than the transmission layer for the generality of synthetic data. Under this assumption, reflection layer after intensity adjustment would be apparently darker than transmission layer.

Fig. 3.5 is the flowchart of data synthesis model.

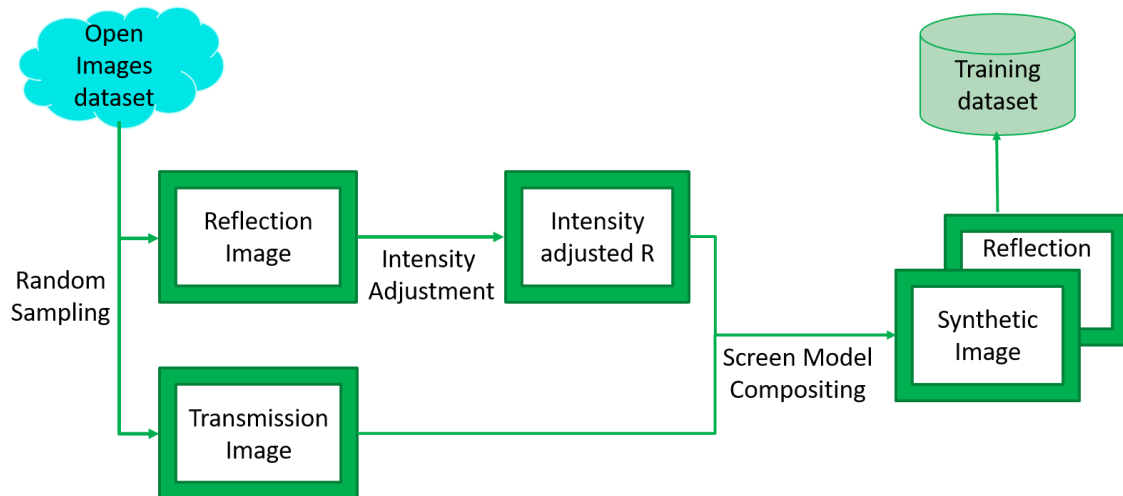


Fig.3.5 Flowchart of data synthesis model.

We use filtered Open Images dataset [19] as our natural image source. Open Images dataset has over 9 million images with 6000 categories provides the diversity of data. First, two images are randomly selected from the dataset and make sure the image pair follow assumption. One of the image would send to intensity adjustment process as reflection layer and the other, as the transmission layer, would be held on for the final compositing process. We use a mapping function as our intensity adjustment. The function is regressed from the real image data capture in different shutter time. Fig. 3.6 shows the regression of data cluster. The intensity of reflection image would be adjusted to 9% to 18% stochastically. Finally, the adjusted reflection image and transmission image would be composited into final image by the screen blend model. A pair of final image and reflection image would be one of our training data. After repeat this process, we can get our training dataset in any desired amount. Fig. 3.7 shows some examples of the synthetic data.

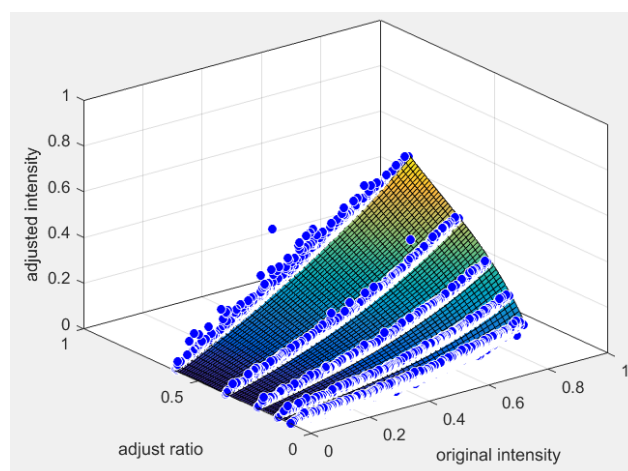


Fig. 3.6 Surface regression of intensity in different energy ratio.

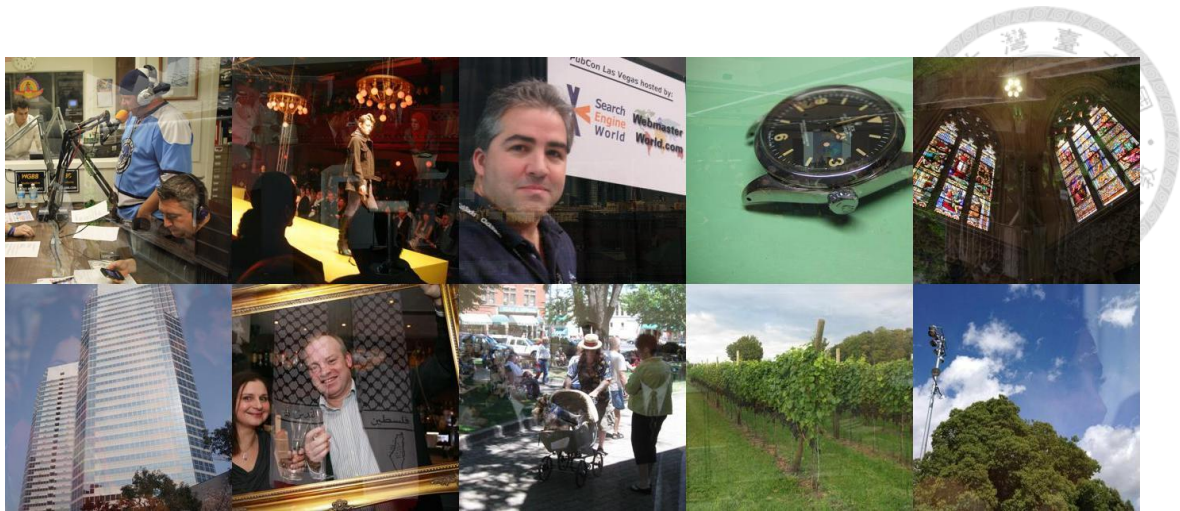


Fig. 3.7 Examples of synthetic data. Best viewed on screen.

3.3 Multi-Task Generative Adversarial Network

Generative Adversarial Network (GAN), proposed by Goodfellow in 2014 [14], has been convinced as a potential generative model especially in generating natural image. In contrast to present generative model as AutoEncoder, Generative Adversarial Network can generate natural image sharper and various. Therefore, many variation of Generative Adversarial Network has been proposed rapidly. Isola et al. [15] proposed a new network based on conditional Generative Adversarial Network to solve the image-to-image translating problem in 2016. The generative model in image-to-image Generative Adversarial Network can generate various image under certain constrains, e.g., edge map. Inspired by the Isola et al., we proposed a variation of Generative Adversarial Network based on image-to-image Generative Adversarial Network architecture. We modified the loss function of generator and import the concept of

multi-task as well to solve the single image reflection removal problem.

In (1), we assume layer compositing as a function f . Our goal is to find the inverse function of f as f' which is approximated by our network. Since we have made assumption to f , it can be our best prior to modify network. Fig. 3.8 is the architecture of image-to-image Generative Adversarial Network. We make two modification. First, L1-norm or L2-norm is most frequently used to be the loss function in the most of the AutoEncoder. As we know, under the screen model assumption, reflection will be suppressed at the high intensity region and remain most at the low intensity region. This significant difference will affect the parameters Generative Adversarial Network learned. The parameters in high intensity region would be very different to the low intensity region. Obviously, we should focus on the low intensity region since it contains stronger reflection but L1-norm or L2-norm treat all pixel equally. Therefore, we used weighted L1-norm to replace the original one. It is defined as follow:

$$WL_1 = (1 - T_{gt}) \odot L_1^R \quad (5)$$

where T_{gt} is the ground truth transmission layer and L_1^R is original L1-norm on generated reflection layer. By using weighted L1-norm, network can pay more effort on the region our vision system care more about.



Fig. 3.8 Architecture of image-to-image Generative Adversarial Network [15].

Second, the high intensity region has inevitably worse performance when we modify the loss function in previous part. In order to solve this problem, we import the concept of multi-task into our network. Specifically, we add another generator into network. Fig. 3.9 is the modified architecture of network. This generator is aim to learn how to separate high intensity region reflection. It has different weight L1-norm from the origin one as follow:

$$WL_1 = T_{gt} \odot L_1^R \quad (6)$$

After two generator give their own result layer, we get our final generating result as follow:

$$R_{final} = (1 - T_l) \odot R_l + T_l \odot R_h \quad (7)$$

where T_l , R_l is transmission layer and reflection layer generated by low region generator respectively, R_h is reflection layer generated by high region generator and R_{final} is the final generating result and the input of discriminator. Fig. 3.10 shows the intermediate output and result of multi-task Generative Adversarial Network and



corresponding PSNR. It shows that result is improved by import multi-task into Generative Adversarial Network.

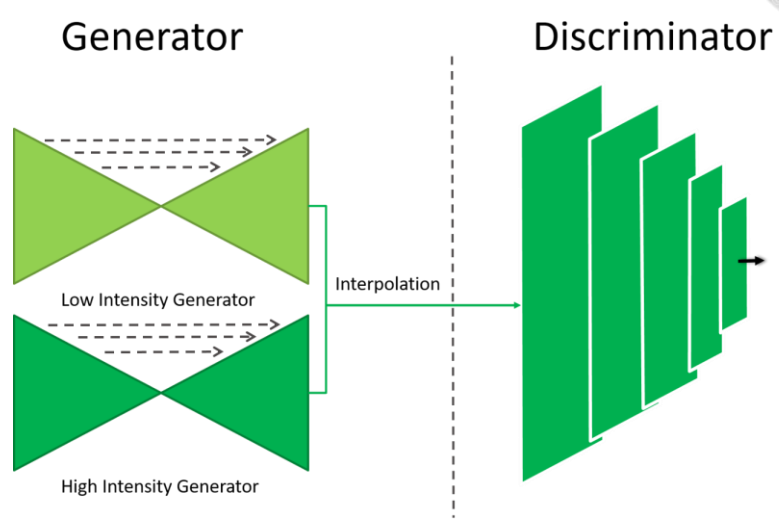


Fig. 3.9 Architecture of Multi-Task Generative Adversarial Network.

Finally, in order to make Generative Adversarial Network more stable to train, we implement Wasserstein GAN with Gradient Penalty [16, 17, 18]. Experiment in [18] shows that WGAN-GP can improve the conventional Generative Adversarial Network which is very hard to train.

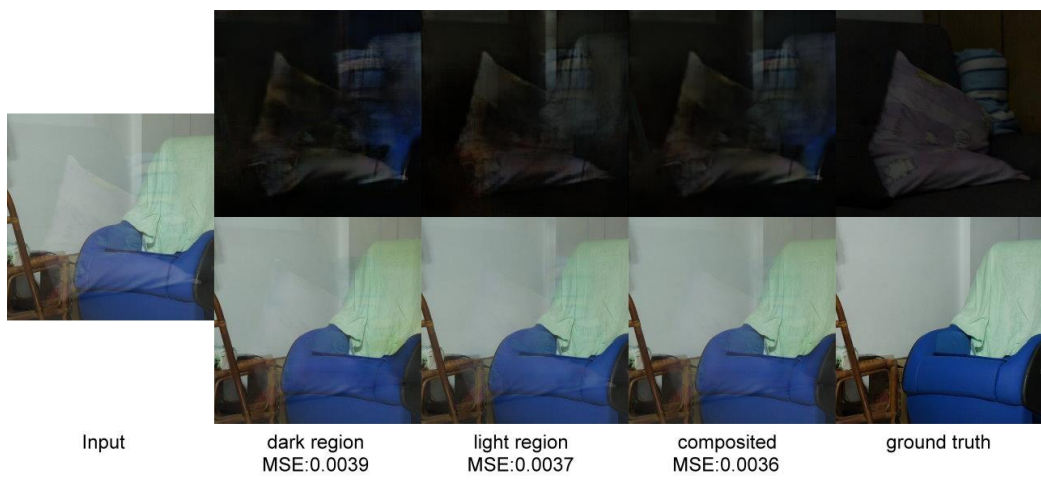


Fig. 3.10 Intermediate output and composited result. Best viewed on screen.

Chapter 4 Experimental Results

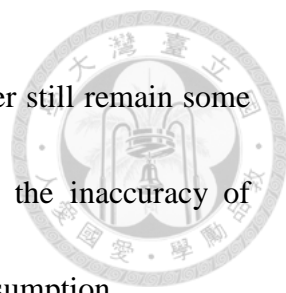


4.1 Experiment Settings

Since we use the base architecture of Image-to-image network, each generator in Generative Adversarial Network is a U-net, which is an AutoEncoder with skip connections and discriminator is PatchGAN in Image-to-image network. We synthesized 100K $256 \text{ pixel} \times 256 \text{ pixel} \times 3 \text{ channel}$ images as training dataset. We gathered some real photos with reflection for evaluation. We train our network on PC with NVIDIA TITAN X GPU. We pre-trained an AutoEncoder with L1-norm loss function. It can provide both of generators a decent initialization and accelerate training progress of Generative Adversarial Network. The batch size is 10 in pre-trained AutoEncoder and multi-task Generative Adversarial Network. Pre-trained AutoEncoder is trained for 50 epochs and multi-task Generative Adversarial Network is 100 epochs. Whole training process takes around 14 days on the previous machine.

4.2 Qualitative Results

In Fig. 4.1, (a) is the input image, (b) is reflection layer of low intensity region, (c) is reflection layer of high intensity region, (d) is composited reflection layer and (e) is corresponding transmission layer. As we can see, reflection layers are all generated



nearly correct in fig. 4.1 though the corresponding transmission layer still remain some reflection. Most artifacts in the transmission layer are caused by the inaccuracy of reflection and the other (the third image) is due to the violation of assumption.

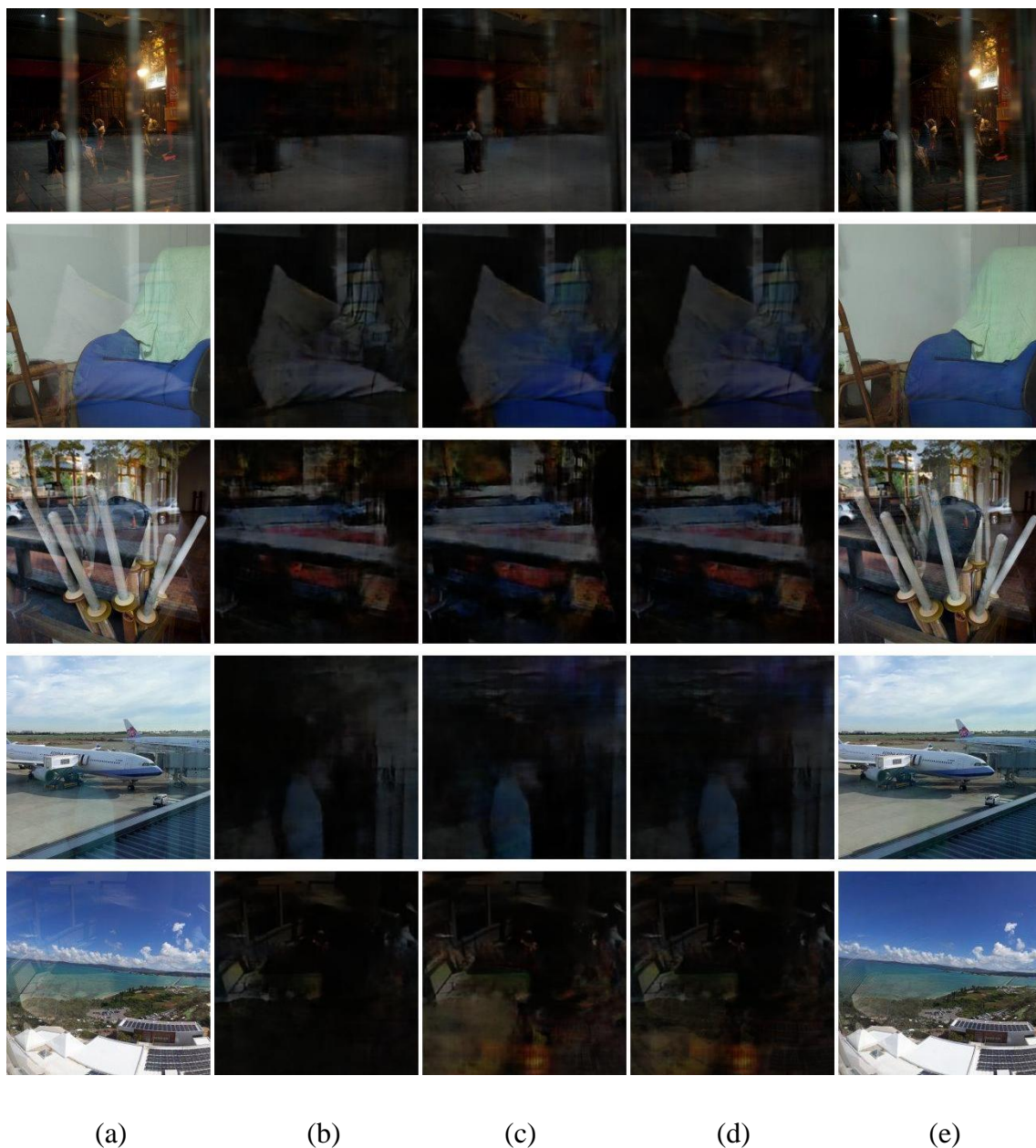


Fig. 4.1 Quality results of successful case. Best viewed on screen.

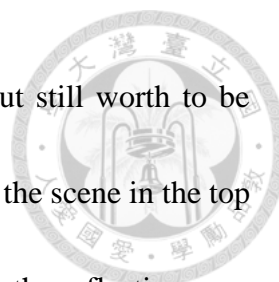


Fig. 4.2 shows some results that can barely be said success but still worth to be mentioned. Although it violated the weakness reflection assumption, the scene in the top two results in the fig. 4.2 is recognizable. The result at the bottom is the reflection over a water surface. Unlike glass, water reflection is more complex due to its waving. These cases attest the robustness of our method.

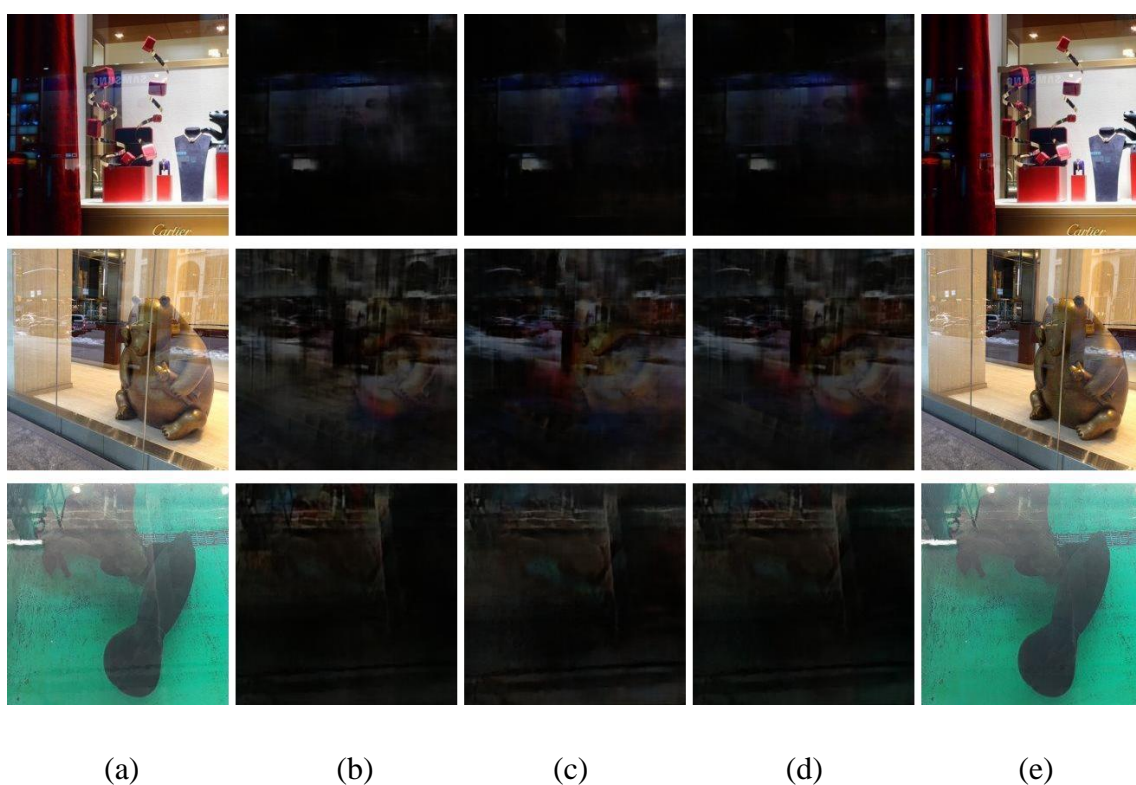
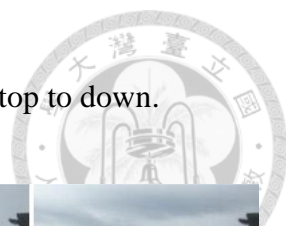


Fig. 4.2 Quality results of some interesting case. Best viewed on screen.

4.3 Compared with other single-image-based approaches

In this section, we are going to compare with three single-image-based approaches: Li and Brown [2], Wan et al. [4] and Arvanitopoulos et al. [5] as state-of-the-art. Fig 4.3



shows the comparing results which reflection strength increase from top to down.

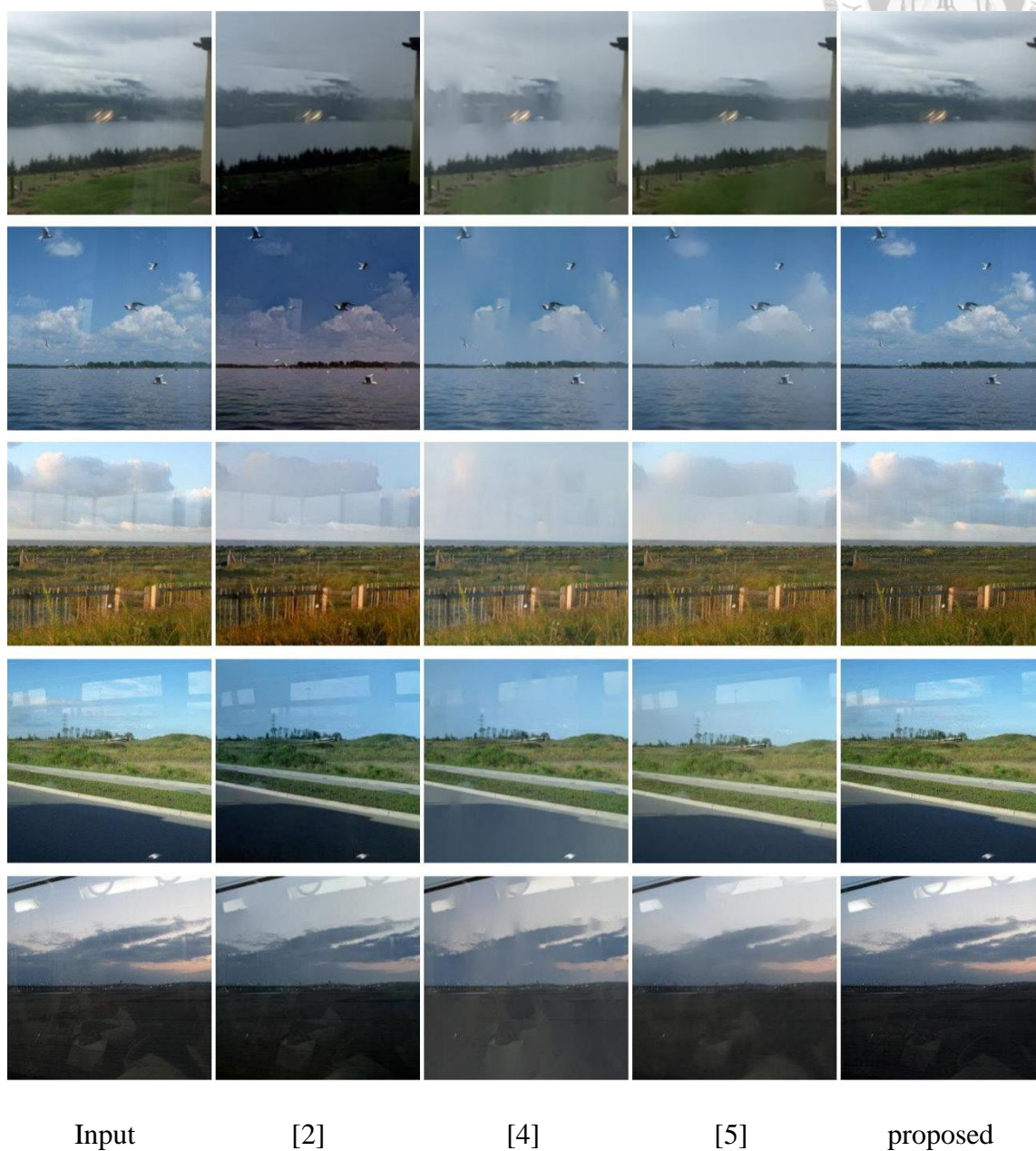


Fig. 4.3 Comparison of other single-image-based approaches. Best viewed on screen.

In first three images, proposed method not only remove reflection layer from the transmission layer but also keep texture and brightness. Li and Brown [2] caused change the tone of the image and remained most of the reflection. Wan et al. [4],



although performed better on reflection removal, lost texture and left obvious artifacts.

Arvanitopoulos et al. [5] smoothed the region that contained reflection so texture and detail were missing significantly. In the last two images, reflection strength is too strong that violated our assumption thus our network failed to remove reflection. However, we found that in the fourth image, our network identified right region of reflection but failed to remove it in one removal due to its high strength. In this kind of case, we can iteratively execute removal process to get better result. Fig. 4.4 shows the all results on four-time removal process.

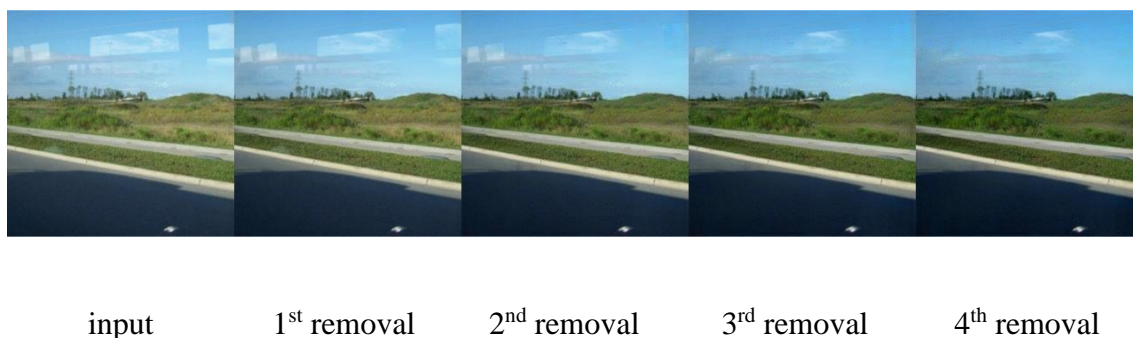


Fig. 4.4 Results of iterative removal on strong reflection. Best viewed on screen.

Chapter 5 Conclusion and Future Work




5.1 Conclusions

Reflection removal from a single image have troubled researchers for a long time. Until today, it is still a highly ill-posed problem waiting us to solved. Conventional approaches utilizing optimization makes it a heavy time-consuming process and unpractical. Existing priors seems to be insufficient to take the solution further. We have to discover new way to break the ice and deep learning based methods could be a good choice. One of the charming characteristic of deep learning based methods on image processing is its ability to dig out completely new features. In this thesis, we proved that multi-task Generative Adversarial Network could cross the wall and make new benchmark for reflection removal. It is not only effective but also fast than conventional approaches. Therefore, to make reflection removal on single image in a reasonable time is no longer a dream.

5.2 Future Work

Considering limited computing power, we limit our input at $256 \text{ pixel} \times 256 \text{ pixel} \times 3$ channel images. However, applicable on high resolution image would be inevitable. In the short term, we can try separating high resolution image into patches



and then removal reflection on patches. Using interpolation or other patch-based merging algorithm to compositing result patches back to high resolution image. In the long term, we should focus on how to apply deep neural network on high dimension input in affordable time and space. Another idea is to refine the data synthesis model. Trying to lose the constraint on reflection strength may help dataset more representative to the population.


REFERENCE



- [1] A. Levin and Y. Weiss. User Assisted Separation of Reflections from a Single Image Using a Sparsity Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(9):1647–1654, 2007.
- [2] Y. Li and M. S. Brown. Single Image Layer Separation Using Relative Smoothness. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2752–2759, 2014.
- [3] Y. Shih, D. Krishnan, F. Durand, and W. T. Freeman. Reflection Removal using Ghosting Cues. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3193– 3201, 2015.
- [4] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot. Depth of field guided reflection removal. In *IEEE International Conference on Image Processing (ICIP)*, pages 21–25, 2016.
- [5] N. Arvanitopoulos, R. Achanta, and S. Süsstrunk. Single Image Reflection Suppression. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [6] Y. Li and M. S. Brown. Exploiting Reflection Change for Automatic Reflection Removal. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2432–2439, 2013.



- [7] X. Guo, X. Cao, and Y. Ma. Robust Separation of Reflection from Multiple Images. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 2195–2202, 2014.
- [8] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman. A Computational Approach for Obstruction-free Photography. *ACM Transactions on Graphics*, 34(4):79:1–79:11, 2015.
- [9] C. Sun, S. Liu, T. Yang, B. Zeng, Z. Wang, and G. Liu. Automatic Reflection Removal Using Gradient Intensity and Motion Cues. In *ACM Multimedia*, MM '16, pages 466–470, 2016.
- [10] J. Yang, H. Li, Y. Dai, and R. T. Tan. Robust Optical Flow Estimation of Double-Layer Images under Transparency or Reflection. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 1410-1419, 2016.
- [11] J. Lai, W. K. Leow, T. Sim, and V. Sharma. Think Big, Solve Small: Scaling Up Robust PCA with Coupled Dictionaries. In *Applications of Computer Vision (WACV)*, pages 1-8, 2016.
- [12] N. Kong, Y. W. Tai, and J. S. Shin. A Physically-Based Approach to Reflection Separation: From Physical Modeling to Constrained Optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI)*, 36(2):209–221, 2014.

- 
- [13] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li. Removing Photography Artifacts Using Gradient Projection and Flash-exposure Sampling. *ACM Transactions on Graphics*, 24(3):828–835, 2005.
- [14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. In *NIPS*, 2014.
- [15] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv preprint arXiv:1611.07004*. 2016
- [16] M. Arjovsky and L. Bottou. Towards Principled Methods for Training Generative Adversarial Networks. *NIPS 2016 Workshop on Adversarial Training*. In review for *ICLR*. Vol. 2016. 2017.
- [17] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*. 2017.
- [18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin,, and A. Courville. Improved Training of Wasserstein GANs. *arXiv preprint arXiv:1704.00028*. 2017.
- [19] I. Krasin, T. Duerig, N. Alldrin, A. Veit, S. Abu-El-Haija, S. Belongie, D. Cai, Z. Feng, V. Ferrari, V. Gomes, A. Gupta, D. Narayanan, C. Sun, G. Chechik, and K. Murphy. OpenImages: A public dataset for large-scale multi-label and multi-class image classification, 2016. Available from <https://github.com/openimages>.