Weather Translation in Images Using Variational Autoencoders

Stefan Stefanache

November 3, 2022

1 Paper Outline

1. Abstract
2. Introduction
    (a) Motivation and Overview
    (b) Contributions
3. Background
    (a) Related Works
4. Our Approach
    (a) Dataset Creation
    (b) Model Architecture
    (c) Training
5. Experiments
    (a) Results
    (b) Comparison with Other Models
    (c) Failure Cases
6. Conclusion
7. References

2 Hypothesis

In this paper, the weather translation task is proposed, which refers to transferring weather conditions of the image from one category to another. It is important for photographic style transfer. This has been attempted before through WeatherGAN [11], where weather conditions are determined by various weather-cues, obtained through attention and segmentation modules, which are then passed to a translation module that generates a new image. Our hypothesis is that variational autoencoders can be trained to give better (more realistic) results in this task, by capturing weather information in their latent space.

3 Methodology

To achieve this result, we follow the steps below:

1. Create dataset containing pictures from the same angle and different weather conditions, labelled with weather categories.

   (a) Scrape images from windy.com webcams from year-round slide show. These contain images with all kinds of weather (foggy, snowy, sunny, cloudy), taken from the same POV (see Fig. 1).

   (b) Drop images of streams having defects like a lot of noise or high variations in scenery (rarely, webcams can be moved to another location).

   (c) Develop a ResNet-based [5] network similar to WeatherNet [7] to label each image with the corresponding weather category.

   Assuming that stream images are from the same physical locations, it's likely that the only differences beteween images is given by the changes in scenery caused by time of the year and weather conditions. This allows autoencoders to "learn" what this difference is and reproduce it.

2. Create and train autoencoders for each possible translation (e.g. sunny $\rightarrow$ cloudy). For the training, we follow the next steps for each ordered pair of images from the same stream in the dataset:

   (a) Based on the labels of the images, choose the coresponding autoencoder. For example, if the images have the sunny and cloudy labels, use the autoencoder that transforms sunny images to cloudy ones.

   (b) Train the chosen encoder to transform from one image to the other e.g. give one image as input of the decoder, and the other as the target output of the decoder.

   This way, we should end up with a set of encoders that can be merged into model that can reliably perform weather translations.

4 Experiments

To check how performant our model is, we'll separate some of our dataset into a test portion, and we'll be be evaluating the model's accuracy on it using 2 metrics traditionally [1] used for quantifying the quality of image generative models:

1. Frechet Inception Distance (FID) [6]: The score summarizes how similar the two groups are in terms of statistics on computer vision features of the raw images calculated using the Inception v3 model used for image classification. Lower scores indicate the two groups of images are more similar, or have more similar statistics, with a perfect score being 0.0 indicating that the two groups of images are identical.

2. Kernel Inception Distance (KID): Calculates the square of the maximum average difference between the two sets of images (input and output), which represents the distribution distance between the two sets of data. In addition, KID has an unbiased estimator, which makes KID close to human perception.

Moreso, this metrics can be used to compare the performance of our model with others like Pix2Pix [8], NVIDIA's UNIT [12] or WeatherGAN [11].

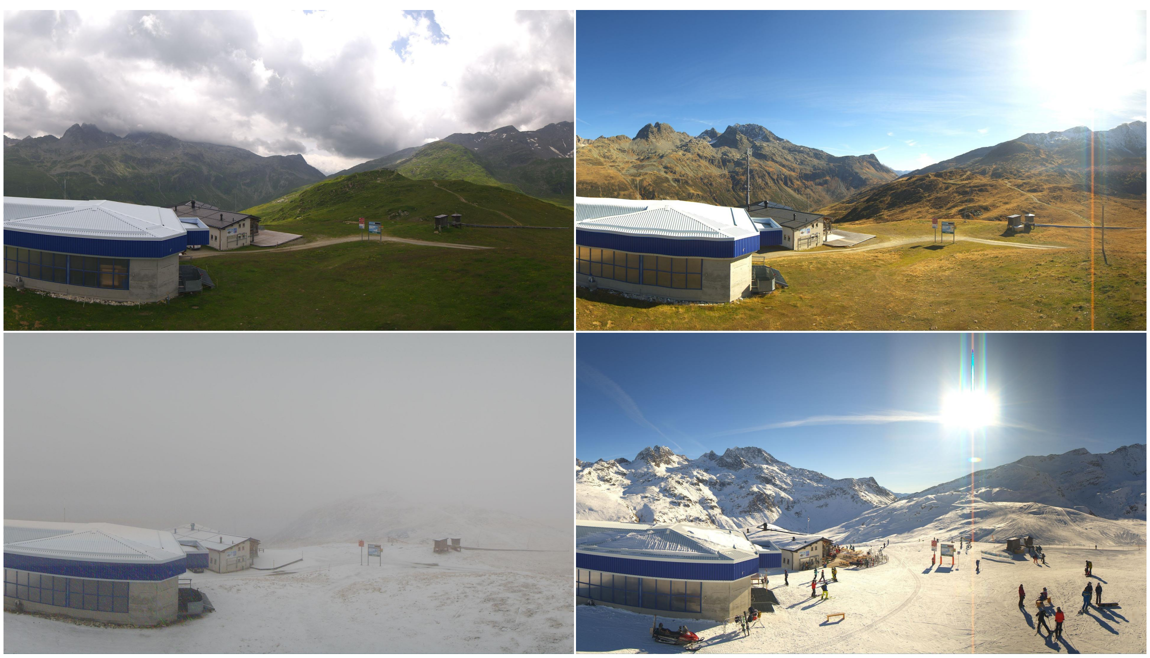

Figure 1: Images from the Splugen Tanatzhöhe Ski Area webcam from windy.com

5 Reference Motivation

1. **ML Concepts & Neural Networks.** Learn more about ML theory and algorithms, along with how Neural Networks are used. Bishop's PRML [2] gives an overview of general ML topics and intodoces Neural Networks in Chapter 5. Goodfellow's [3] book gives an introductive overview of Deep Learning. Specific Neural Networks architectures like ResNet [5] or GANs [4] are useful for understanding previous weather recognition and image translation approaches.

2. **Weather Recogniton using ResNet.** Study previous approaches [7, 13, 11] of recognizing weather conditions in images using different ResNet architectures. Either try to use/implement one of the papers or come up with a slight variation.

3. **Variational Encoders.** Learn about Variational Encoders and the theory behind them: original paper [10], introductive overview [9].

4. **Image Translation.** Study how image translation is done, from generic frameworks like Pix2Pix [8] or NVIDIA's UNIT [12], to the weather specific case - WeatherGAN [11].

6 Selected References

6.1 Deep Residual Learning for Image Recognition [5]

- **Summary:** Offers an overview of residual learning and the ResNet architecture. Useful for recognizing weather conditions in images and constructing the dataset.
- **138531 Citations** and **49 References** (IEEE style)
- **Chapters:**
  1. Introduction
  2. Related Work
  3. Deep Residual Learning
  4. Experiments
  5. References

6.2 WeatherNet: Recognising weather and visual conditions from street-level [7]

- **Summary:** A pipeline of four deep Convolutional Neural Network (CNN) models, so-called the WeatherNet, is trained, relying on residual learning using ResNet50 architecture, to extract various weather and visual conditions such as Dawn/dusk, day and night for time detection, and glare for lighting conditions, and clear, rainy, snowy, and foggy for weather conditions.

- **33 Citations** and **39 References** (IEEE style)
- **Chapters:**
  1. Introduction
  2. Related Work
  3. WeatherNet Frameworks
  4. Results
  5. Discussion
  6. Remarks and Future Work
  7. Acknowledgement
  8. References

6.3 An Introduction to Variational Autoencoders [10]

- **Summary:** Variational autoencoders provide a principled framework for learning deep latent-variable models and corresponding inference models. In the paper, an introduction to variational autoencoders and some important extensions is provided
- **1077 Citations** and **213 References** (IEEE style)
- **Chapters:**
  1. Introduction
  2. Variational Autoencoders
  3. Beyond Gaussian Posteriors
  4. Deeper Generative Models
  5. Conclusion
  6. Acknowledgement
  7. References

6.4 Unsupervised Image-to-Image Translation Networks [12]

- **Summary:** Unsupervised image-to-image translation aims at learning a joint distribution of images in different domains by using images from the marginal distributions in individual domains. Since there exists an infinite set of joint distributions that can arrive the given marginal distributions, one could infer nothing about the joint distribution from the marginal distributions without additional assumptions. To address the problem, a shared-latent space assumption is made and an unsupervised image-to-image translation framework based on Coupled GANs is proposed.
- **2359 Citations** and **29 References** (IEEE style)
- **Chapters:**
  1. Introduction
  2. Assumptions
  3. Framework
  4. Experiments
  5. Related Work
  6. Conclusion and Future Work
  7. Network Architecture
  8. Domain Adaptation

6.5 Weather GAN: Multi-Domain Weather Translation Using Generative Adversarial Networks [11]

- **Summary:** A new task is proposed, namely, weather translation, which refers to transferring weather conditions of the image from one category to another. It is important for photographic style transfer. Although lots of approaches have been proposed in traditional image translation tasks, few of them can handle the multi-category weather translation task, since weather conditions have rich categories and highly complex semantic structures. To address this problem, a multi-domain weather translation approach based on generative adversarial networks (GAN) is developed, denoted as Weather GAN, which can achieve the transferring of weather conditions among sunny, cloudy, foggy, rainy and snowy. Specifically, the weather conditions in the image are determined by various weather-cues, such as cloud, blue sky, wet ground, etc.
- **9 Citations** and **50 References** (IEEE style)
- **Chapters:**

1. Introduction
2. Related Work
3. Our Approachj
4. Experiments
5. Conclusion
6. References

7 Paper Classification

1. **ACM**
   (a) I.2.10: Vision and Scene Understanding
   (b) I.4.8: Scene Analysis
2. **AMS**
   (a) 68T07. Artificial neural networks and deep learning
   (b) 68T45. Machine vision and scene understanding

References

[1] Eyal Betzalel, Coby Penso, Aviv Navon, and Ethan Fetaya. A study on the evaluation of generative models, 2022. URL: `https://arxiv.org/abs/2206.10935`, `doi:10.48550/ARXIV.2206.10935`.

[2] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. URL: `https://arxiv.org/abs/1406.2661`, `doi:10.48550/ARXIV.1406.2661`.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL: `http://arxiv.org/abs/1512.03385`, `arXiv:1512.03385`.

[6] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. 2017. URL: `https://arxiv.org/abs/1706.08500`, `doi:10.48550/ARXIV.1706.08500`.

[7] Mohamed R. Ibrahim, James Haworth, and Tao Cheng. WeatherNet: Recognising weather and visual conditions from street-level images using deep residual learning. *CoRR*, abs/1910.09910, 2019. URL: http://arxiv.org/abs/1910.09910, arXiv:1910.09910.

[8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. URL: http://arxiv.org/abs/1611.07004, arXiv:1611.07004.

[9] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013. URL: https://arxiv.org/abs/1312.6114, doi:10.48550/ARXIV.1312.6114.

[10] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *CoRR*, abs/1906.02691, 2019. URL: http://arxiv.org/abs/1906.02691, arXiv:1906.02691.

[11] Xuelong Li, Kai Kou, and Bin Zhao. Weather GAN: multi-domain weather translation using generative adversarial networks. *CoRR*, abs/2103.05422, 2021. URL: https://arxiv.org/abs/2103.05422, arXiv:2103.05422.

[12] Ming-Yu Liu, Thomas M. Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *CoRR*, abs/1703.00848, 2017. URL: http://arxiv.org/abs/1703.00848, arXiv:1703.00848.

[13] Bin Zhao, Xuelong Li, Xiaoqiang Lu, and Zhigang Wang. A CNN-RNN architecture for multi-label weather recognition. *CoRR*, abs/1904.10709, 2019. URL: http://arxiv.org/abs/1904.10709, arXiv:1904.10709.