Weather Translation in Images Using Variational Autoencoders

Stefan Stefanache

June 14, 2023

1 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 [14], 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 convolutional variational autoencoders can be trained to give better (more realistic) results in this task, by using a specialized dataset containing paired (from the same POV) images with different weather conditions. **Disclaimer:** Currently, the paper describes the dataset and its construction, specifically for the sunny/cloudy cases, along with a draft proposal for the VAE model architecture. The code is available at github.com/thesstefan/vae_weather_translation. and the dataset is available here.

2 Related Work

In this section, we introduce some state-of-the art approaches of weather recognition and image-to-image translation, as well as a brief discussion on datasets created for various tasks related to weather.

2.1 Weather Recognition

Elhoseiny et al. [4] tackles the two-class weather classification task by the AlexNet [13], which has got superior performance than handcrafted features. Moreover, to combine the advantages of deep feature and handcrafted feature, these two kinds of features are integrated in [16]. The multi-task model proposed by Zhao et al. [22] identifies weather cues in images through segmentation and then uses them to classify images, achieving state-of-the-art performance.

2.2 Weather Datasets

One popular dataset for weather recognition is called **MWI** [20, 21] (Multiclass Weather Image). It contains 20K images obtained from many web albums and films, such as Flicker, Picasa, Poco, Fengniao. The images are collected by several helpers, and they choose images with their own common sense. The main purpose of this dataset is to provide an extensive testbed for the evaluation of existing appearance models, and provide insight needed to develop new appearance models.

To facilitate weather property estimation from images, a large-scale image dataset associated with rich weather information called **Image2Weather** [3] was developed. Based on the taken time and geographical information of an image, weather properties obtained from a weather forecast website are linked with the images. Through data filtering like indoor/outdoor classification and sky region detection, the dataset consisting of more than 180,000 photos is built to promote related research.

2.3 Image-To-Image Translation

Studies have explored various models with GANs for image-to-image translation. For example, Pix2pix, used a conditional, generative adversarial network [17] for image translation, when a source and target image pair was given. Similar methods have been adopted for several tasks, such as synthesizing a photograph from a sketch [18]. Weather translation was also approached using GANs, through WeatherNet [9].

When paired training data were not available for some tasks, other studies have suggested several approaches to transform from an unpaired image. This has been approached from both the perspective of coupled GANs (UNIT [15]), and by using cross-domain variational encoders and decoders [19].

3 Our Approach

In this paper, we propose an image-to-image translation framework for the weather translation task, using CVAEs (convolutional variational autoencoders) powered by a specialized dataset (called **weather2weather**), created for this purpose.

3.1 The weather2weather Dataset

The **weather2weather** dataset is constructed from scratch by ourselves. It consists of **TBD** (currently around 5000) images scraped from windy.com, a provider of webcam streams. Images are separated by the webcam streams they belong from, so one can create pairs of images from the same location, which can then be used for paired image-to-image translation tasks related to weather. Weather-cues for each image are extracted in a segmentation mask using the multi-task weather recognition model proposed in [22], which are then used to assign a weather label to the image. The process of constructing the dataset is illustrated in 1. Stream images and their inferred properties are also cleaned and verified by the authors, who check for misclassifications or unsuitable images using tools like the one shown in 2.

Currently, the model only classifies images into sunny or cloudy. The model could also be extended for multiclass classification for other cases like foggy, snowy or rainy weather. The process of enhancing the **weather2weather** dataset could be automated further by handling some edge cases, like detecting noisy images or webcams that are moving.

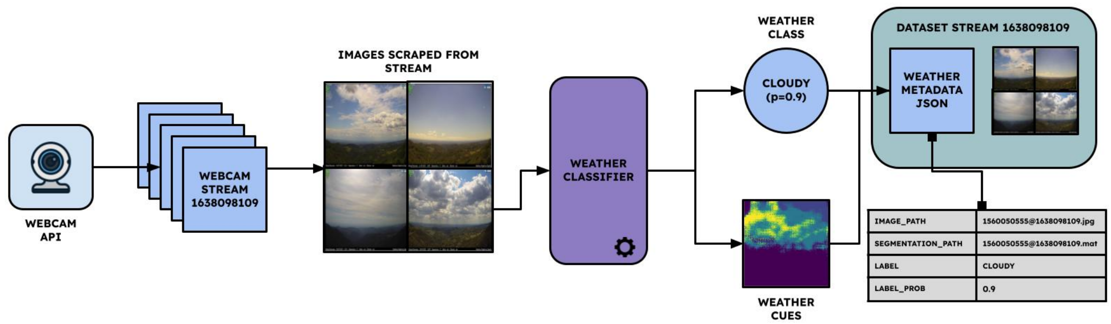

Figure 1: Creating the **weather2weather** dataset

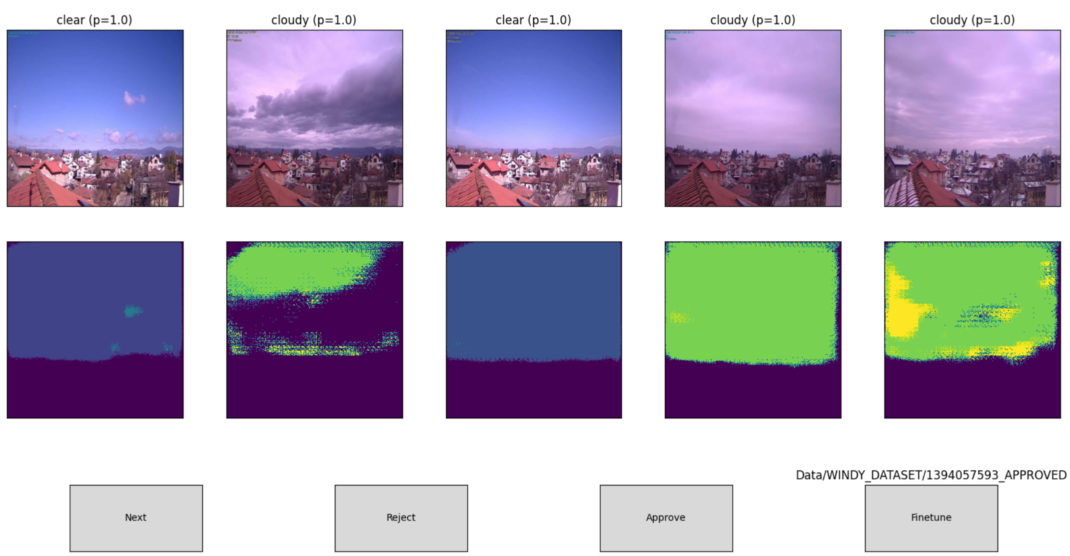

Figure 2: Validating streams in the **weather2weather** dataset

3.2 Weather-To-Weather Translation Model

Disclaimer: Exact model architecture and training details **TBD**.

Convolutional variational auto-encoders are used to learn about the distribution of images and weather cues (segmentation masks) for each weather category (currently cloudy or sunny). The bottlenecks of both decoders are then merged through a composition (**TBD**) operation. The model samples a data sample from the resulting distribution, which then is taken through a decoder that produces the translated image. Note that each decoder is also trained on a corresponding weather category. Generally, the decoders and encoders used are chosen based on the input image weather category and the target weather category.

The train data is obtained by taking each pair of images with different weather categories from every stream in the **weather2weather** dataset, resulting in over 15000 such pairs. Train images and segmentations masks are pulled through the model. To train the auto-encoders and decoders for each domain, we follow the basic equation of the convolutional auto-encoder [11]. The representation of the k-th latent feature map extraction for input I is denoted as

$$h^k = \sigma(I * W^k + b^k) \tag{1}$$

where σ is an activation function, W_k is a weight, b^k is a bias, and $*$ represents the 2D convolution operation. The reconstruction of the latent representation is obtained using

$$y = \sigma\left(\int_{k \in H} h^k * \widehat{W} + c \right) \tag{2}$$

where c is a bias for input channel. The model is trained to minimize the ELBO loss [11], which is defined as

$$\min \mathbb{E}_q[\log q(z|x) - \log p(z)] - \mathbb{E}_q \log p(x|z) \tag{3}$$

where the first term is the Kullback-Leibler divergence and the second term is the construction loss.

The first step of translation is taking an image and inferring its weather cues and weather label using the model proposed by [22]. Based on the classification label, the corresponding encoders are used to compress the image and its segmentation mask to produce the merged latent space. A sample then is taken from the merged sample space to produce the translated image using the corresponding encoder. The process is illustrated in 3.

4 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.

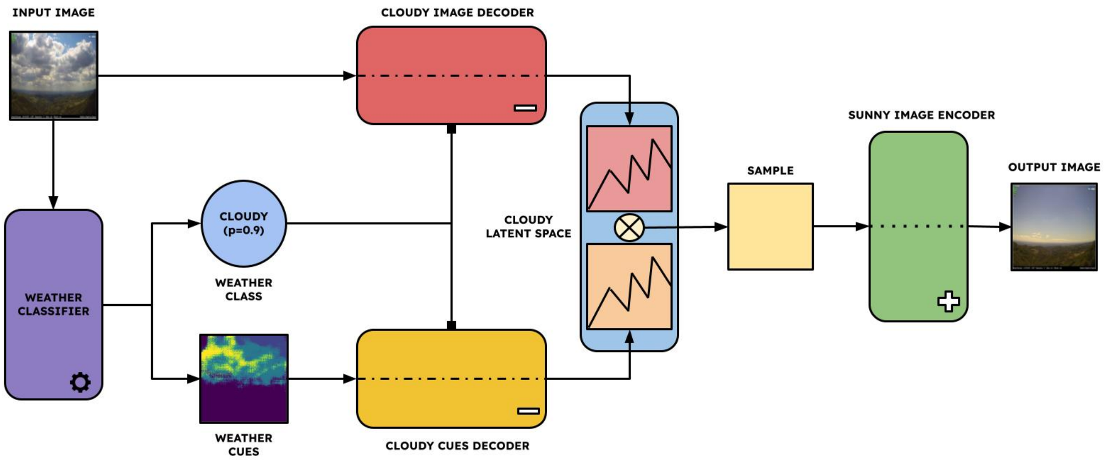

Figure 3: The weather translation process

(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. 4).

(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 [7] network similar to WeatherNet [9] 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.

5 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) [8]: 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 [10], NVIDIA's UNIT [15] or WeatherGAN [14].

6 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 intoduces Neural Networks in Chapter 5. Goodfellow's [5] book gives an introductive overview of Deep Learning. Specific Neural Networks architectures like ResNet [7] or GANs [6] are useful for understanding previous weather recognition and image translation approaches.

2. **Weather Recogniton using ResNet.** Study previous approaches [9, 23, 14] 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 [12], introductive overview [11].

4. **Image Translation.** Study how image translation is done, from generic frameworks like Pix2Pix [10] or NVIDIA's UNIT [15], to the weather specific case - WeatherGAN [14].

Figure 4: Images from the Splugen Tanatzhöhe Ski Area webcam from windy.com

7 Selected References

7.1 Deep Residual Learning for Image Recognition [7]

- **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

7.2 WeatherNet: Recognising weather and visual conditions from street-level [9]

- **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

7.3 An Introduction to Variational Autoencoders [12]

- **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

7.4 Unsupervised Image-to-Image Translation Networks [15]

- **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

7.5 Weather GAN: Multi-Domain Weather Translation Using Generative Adversarial Networks [14]

- **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

8 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

9 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

References

[1] Eyal Betzalel, Coby Penso, Aviv Navon, and Ethan Fetaya. A study on the evaluation of generative models, 2022.

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

[3] Wei-Ta Chu, Xiang-You Zheng, and Ding-Shiuan Ding. Image2weather: A large-scale image dataset for weather property estimation. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, pages 137–144, 2016.

[4] Mohamed Elhoseiny, Sheng Huang, and Ahmed Elgammal. Weather classification with deep convolutional neural networks. pages 3349–3353, 2015.

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

[6] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[8] 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.

[9] 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.

[10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[11] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013.

[12] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *CoRR*, abs/1906.02691, 2019.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*.

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

[15] Ming-Yu Liu, Thomas M. Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *CoRR*, abs/1703.00848, 2017.

[16] Cewu Lu, Di Lin, Jiaya Jia, and Chi-Keung Tang. Two-class weather classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2510–2524, 2017.

[17] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

[18] Patsorn Sangkloy, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Scribbler: Controlling deep image synthesis with sketch and color. *CoRR*, abs/1612.00835, 2016.

[19] Jaechang Yoo, Heesong Eom, and Yong Choi. Image-to-image translation using a cross-domain auto-encoder and decoder. *Applied Sciences*, 9:4780, 11 2019.

[20] Zheng Zhang and Huadong Ma. Multi-class weather classification on single images. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4396–4400, 2015.

[21] Zheng Zhang, Huadong Ma, Huiyuan Fu, and Cheng Zhang. Scene-free multi-class weather classification on single images. *Neurocomputing*, 207:365–373, 2016.

[22] Bin Zhao, Lulu Hua, Xuelong Li, Xiaoqiang Lu, and Zhigang Wang. Weather recognition via classification labels and weather-cue maps. *Pattern Recognition*, 95:272–284, 2019.

[23] Bin Zhao, Xuelong Li, Xiaoqiang Lu, and Zhigang Wang. A CNN-RNN architecture for multi-label weather recognition. *CoRR*, abs/1904.10709, 2019.