Photographic Visualization of Weather Forecasts with Generative Adversarial Networks

ECMWF Machine Learning Workshop
March 31st, 2022

Christian Sigg (MeteoSwiss), Flavia Cavallaro (Comerge), Tobias Günther (FAU) and Martin R. Oswald (ETH Zürich and UvA)
Contact: christian.sigg@meteoswiss.ch
Outline

Motivation: Why Photographic Images?
Baseline and Evaluation Criteria
Method: Conditional GANs
Results
Conclusions and Future Work
Outdoor Weather Cameras

An information-dense yet accessible visualization of past and present weather:
Visualization of Weather Forecasts

Screenshots of the MeteoSwiss smartphone app

Also use photographic images to visualize future weather conditions!
Outline

- Motivation: Why Photographic Images?
- Baseline and Evaluation Criteria
- Method: Conditional GANs
- Results
- Conclusions and Future Work
Baseline: Analog Retrieval

$\hat{I}_t$ Image sequence taken at Flüela, 10 to 16 UTC on July 2$^{nd}$, 2020

Retrieval of best matching individual images from annotated archive

Retrieval of best matching sequence
I. Images should look real, no obvious artifacts
II. Match future atmospheric, ground and illumination conditions
III. Seamless transition from observation to forecast
IV. Visual continuity between consecutive images
Evaluation of Analog Retrieval

<table>
<thead>
<tr>
<th>I. Realism</th>
<th>II. Matching conditions</th>
<th>III. Seamless transition</th>
<th>IV. Visual continuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analog images</td>
<td>😊</td>
<td>😞</td>
<td>😞</td>
</tr>
<tr>
<td>Analog sequence</td>
<td>😊</td>
<td>😞</td>
<td>😞</td>
</tr>
</tbody>
</table>

High information density of images → retrieving analogs is not feasible 😞
Outline

- Motivation: Why Photographic Images?
- Baseline and Evaluation Criteria
- Method: Conditional GANs
- Results
- Conclusions and Future Work
Image Synthesis: A Regression Problem

Generate photographic image $\hat{I}_t$, given forecast $w_t$ of future weather conditions

$$G: w_t \mapsto \hat{I}_t$$

Generator $G(w; \theta)$ is a neural network, $\theta$ trained by minimizing expected loss

$$\arg\min_{\theta} \mathbb{E}_{w_t, I_t} [L(G(w; \theta), I_t)]$$
Choice of Loss Function $L$

$$\arg\min_\theta \mathbb{E}_{w_t,I_t} [L(G(w_t; \theta), I_t)]$$

Forecast $w_t$ does not determine exact shapes and locations of clouds → Pixel-wise loss function is not appropriate, results in uniform sky:

Goal: User should not be able to tell whether $I_t$ or $\hat{I}_t$ is the real image, even if they are not identical.
Generative Adversarial Networks  Goodfellow et al., 2014

**Discriminator** $D: I \mapsto [0, 1]$ mimics user, learns loss function through adversarial training

**Generator** $G: z \mapsto I$, creates image $I$ from random input $z \sim \mathcal{N}(0, 1)$

$$\min_{\theta} \max_{\eta} \mathbb{E}_I [\log D(I; \eta)] + \mathbb{E}_z [\log(1 - D(G(z; \theta); \eta))]$$

- **authenticate real images**
- **fool discriminator**
- **spot fake images**
Generator Architecture

- **Conditional Generator** [Mirza and Osindero, 2014] transforms current image $I_0$
- **Encoder-decoder with skip connections** [Ronneberger et al., 2015]
- **Spectral normalization applied to all convolution layers** [Miyato et al., 2018]
Discriminator Architecture

- Conditional discriminator $D(I|I_0, w_0, w_t)$
- Two output heads: patch-level $D_p$ and pixel-l $D_{ij}$  

Schonfeld et al., 2020
Motivation: Why Photographic Images?
Baseline and Evaluation Criteria
Method: Conditional GANs
Results
Conclusions and Future Work
Evaluation Data

Descriptor $w$: time of day, day of year, 31 COSMO-1 hourly output fields

Training: all pairs $(I_0, w_0)$ and $(I_t, w_t)$, $t \in [0, 10, 20, \ldots, 360 \text{ min}]$ of 2019

Test: Jan to Aug of year 2020 (until decommissioning of COSMO-1 at MCH)

Downscaled to 64 x 128 pixels to speed up training and conserve GPU memory
I. Realism

What is your first impression of the image?

- generated
- real
- generated
- real
Results of study with 5 professional users of MCH camera feeds:

<table>
<thead>
<tr>
<th>Actual</th>
<th>Judgment</th>
<th>Actual</th>
<th>Judgment</th>
<th>Actual</th>
<th>Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>57</td>
<td>Real</td>
<td>52</td>
<td>Real</td>
<td>57</td>
</tr>
<tr>
<td>Generated</td>
<td>43</td>
<td>Generated</td>
<td>32</td>
<td>Generated</td>
<td>18</td>
</tr>
</tbody>
</table>

| Generated | 32 | Generated | 43 |

- **Cevio**: 59% accuracy
- **Etziken**: 63% accuracy
- **Flüela**: 55% accuracy

User accuracy is not much better than random guessing 😊
II. Matching Weather Conditions

Atmosphere: cloud cover, cloud type, visibility
Ground: dry, wet, frost, snow
Illumination: time of day, diffuse or direct
II. Matching Weather Conditions

<table>
<thead>
<tr>
<th>Camera</th>
<th>Cloud cover</th>
<th>Cloud type</th>
<th>Visibility</th>
<th>Ground</th>
<th>Time of day</th>
<th>Diffuse/direct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cevio</td>
<td>32</td>
<td>35</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>Etziken</td>
<td>36</td>
<td>36</td>
<td>44</td>
<td>45</td>
<td>45</td>
<td>38</td>
</tr>
<tr>
<td>Flüela</td>
<td>31</td>
<td>33</td>
<td>26</td>
<td>44</td>
<td>41</td>
<td>35</td>
</tr>
</tbody>
</table>

Example: Mismatch in cloud cover

but forecast $w_t$ predicted 100 % cloud area fraction in medium troposphere
## II. Matching Weather Conditions

<table>
<thead>
<tr>
<th>Camera</th>
<th>Cloud cover</th>
<th>Cloud type</th>
<th>Visibility</th>
<th>Ground</th>
<th>Time of day</th>
<th>Diffuse/direct</th>
<th>Viz. failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cevio</td>
<td>32</td>
<td>35</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>Etziken</td>
<td>36</td>
<td>36</td>
<td>44</td>
<td>45</td>
<td>45</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>Flüela</td>
<td>31</td>
<td>33</td>
<td>26</td>
<td>44</td>
<td>41</td>
<td>35</td>
<td>5</td>
</tr>
</tbody>
</table>

**Visualization failure:** forecast $w_t$ is accurate, but generated image $\hat{I}_t$ is inconsistent with it
Possible because $G$ is conditioned on $I_0$, compare to analog retrieval:
Outline

- Motivation: Why Photographic Images?
- Baseline and Evaluation Criteria
- Method: Conditional GANs
- Results
- Conclusions and Future Work
Conclusions

- Photographic images can also visualize future weather conditions
- Look realistic, match predicted weather conditions, attain seamless transition from observation to forecast and visual continuity

Applications:
- Communicate localized forecasts in webcam feeds, smartphone app
- Provide similar service to communities and tourism organizations
Future Work

- Use more accurate and descriptive weather descriptors
- Scale image size beyond 64 x 128 pixels e.g. Karras et al., 2018
- Improve transformations involving translations:

\[
I_t \quad \hat{I}_t
\]

- \( t = 0 \) \quad \( t = 1 \text{ h} \) \quad \( t = 2 \text{ h} \) \quad \( t = 3 \text{ h} \) \quad \( t = 4 \text{ h} \) \quad \( t = 5 \text{ h} \) \quad \( t = 6 \text{ h} \)

(including self-attention layers Zhang et al., 2019 did not help)
- Synthesize whole sequences to improve temporal evolution Wu et al., 2020
The pre-print of our paper is available at

https://arxiv.org/abs/2203.15601

Tensorflow code, trained models and results are available at

https://github.com/meteoswiss/photocast
We thank Rega for giving us the permission to use images from the Cevio camera in this study.

We thank Tanja Weusthoff for the preparation of the COSMO-1 forecast data.

We thank Christian Allemann, Yannick Bernard, Eliane Thürig, Deborah van Geijtenbeek and Abbès Zerdouk for evaluating the realism of individual generated images.

We thank Daniele Nerini for providing his expertise on nowcasting and post-processing of forecasts.
Bibliography


Generator Objectives to be Minimized

How much $G(I_0, z|w_0, w_t)$ struggles to fool the discriminator on the patch level

$$\mathbb{E}_{I_0, w_0, w_t} \mathbb{E}_z \left[ \sum_p \log[D_p(G(I_0, z|w_0, w_t)|I_0, w_0, w_t)] \right]$$

and on the pixel level

$$\mathbb{E}_{I_0, w_0, w_t} \mathbb{E}_z \left[ \sum_{ij} \log[D_{ij}(G(I_0, z|w_0, w_t)|I_0, w_0, w_t)] \right]$$

How similar two generated images look at the pixel level, given different random inputs $z_1, z_2 \sim \mathcal{N}(0, 1)$

$$-\mathbb{E}_{I_0, w_0, w_t} \mathbb{E}_{z_1, z_2} \left[ \sum_{ijc} \left| G_{ijc}(I_0, z_1|w_0, w_t) - G_{ijc}(I_0, z_2|w_0, w_t) \right| \right]$$
Discriminator Objectives to be Maximized

How well the patch head $D_p$ authenticates real images

$$\mathbb{E}_{I_0,w_0,I_t,w_t} \left[ \sum_p \log D_p(I_t|I_0,w_0,w_t) \right]$$

and spots generated images

$$\mathbb{E}_{I_0,w_0,w_t} \mathbb{E}_Z \left[ \sum_p \log[1 - D_p(G(I_0,z|w_0,w_t)|I_0,w_0,w_t)] \right]$$

How well the pixel head $D_{ij}$ can distinguish pixels of a cut-mix composite $C$

$$\mathbb{E}_C \left[ \sum_{ij} M_{ij} D_{ij}(C) + (1 - M_{ij}) \log(1 - D_{ij}(C)) \right]$$

Yun et al., 2019
Artifacts Induced by Residual Learning  He et al., 2015

Clouds in $I_0$ are still partially visible in the clear sky regions of $\hat{I}_t$
→ Residual transformation learned by the generator does not fully cancel their appearance
## Subset of COSMO-1 Output Fields

Schättler et al., 2021

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Unit</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALB_RAD</td>
<td>%</td>
<td>Surface albedo for visible range, diffuse</td>
</tr>
<tr>
<td>ASOB_S</td>
<td>W/m²</td>
<td>Net short-wave radiation flux at surface</td>
</tr>
<tr>
<td>ASWDIFD_S</td>
<td>W/m²</td>
<td>Diffuse downward short-wave radiation at the surface</td>
</tr>
<tr>
<td>ASWDIFU_S</td>
<td>W/m²</td>
<td>Diffuse upward short-wave radiation at the surface</td>
</tr>
<tr>
<td>ASWDIR_S</td>
<td>W/m²</td>
<td>Direct downward short-wave radiation at the surface</td>
</tr>
<tr>
<td>ATHB_S</td>
<td>W/m²</td>
<td>Net long-wave radiation flux at surface</td>
</tr>
<tr>
<td>CLCH</td>
<td>%</td>
<td>Cloud area fraction in high troposphere (pressure below ca. 400 hPa)</td>
</tr>
<tr>
<td>CLCM</td>
<td>%</td>
<td>Cloud area fraction in medium troposphere (between ca. 400 and 800 hPa)</td>
</tr>
<tr>
<td>CLCL</td>
<td>%</td>
<td>Cloud area fraction in low troposphere (pressure above ca. 800 hPa)</td>
</tr>
<tr>
<td>CLCT</td>
<td>%</td>
<td>Total cloud area fraction</td>
</tr>
<tr>
<td>D.TD.2M</td>
<td>K</td>
<td>2 m dew point depression</td>
</tr>
<tr>
<td>DD.10M</td>
<td>°</td>
<td>10 m wind direction</td>
</tr>
<tr>
<td>DURSUN</td>
<td>s</td>
<td>Duration of sunshine</td>
</tr>
<tr>
<td>FF.10M</td>
<td>m/s</td>
<td>10 m wind speed</td>
</tr>
<tr>
<td>GLOB</td>
<td>W/m²</td>
<td>Downward shortwave radiation flux at surface</td>
</tr>
<tr>
<td>H.SNOW</td>
<td>m</td>
<td>Snow depth</td>
</tr>
<tr>
<td>HPBL</td>
<td>m</td>
<td>Height of the planetary boundary layer</td>
</tr>
<tr>
<td>PS</td>
<td>Pa</td>
<td>Surface pressure (not reduced)</td>
</tr>
<tr>
<td>RELHUM.2M</td>
<td>%</td>
<td>2 m relative humidity (with respect to water)</td>
</tr>
<tr>
<td>T.2M</td>
<td>K</td>
<td>2 m air temperature</td>
</tr>
<tr>
<td>TD.2M</td>
<td>K</td>
<td>2 m dew point temperature</td>
</tr>
<tr>
<td>TOT_PREC</td>
<td>kg/m²</td>
<td>Total precipitation</td>
</tr>
<tr>
<td>TOT.RAIN</td>
<td>kg/m²</td>
<td>Total precipitation in rain</td>
</tr>
<tr>
<td>TOT.SNOW</td>
<td>kg/m²</td>
<td>Total precipitation in snow</td>
</tr>
<tr>
<td>U.10M</td>
<td>m/s</td>
<td>10 m grid eastward wind</td>
</tr>
<tr>
<td>V.10M</td>
<td>m/s</td>
<td>10 m grid northward wind</td>
</tr>
<tr>
<td>VMAX.10M</td>
<td>m/s</td>
<td>Maximum 10 m wind speed</td>
</tr>
</tbody>
</table>