

Could artificial intelligence win the next Weather Photographer of the Year competition?

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Hunt, K. M. R. ORCID: <https://orcid.org/0000-0003-1480-3755>
(2022) Could artificial intelligence win the next Weather
Photographer of the Year competition? *Weather*. ISSN 0043-
1656 doi: <https://doi.org/10.1002/wea.4348> Available at
<https://centaur.reading.ac.uk/109017/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/wea.4348>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Could artificial intelligence win the next Weather Photographer of the Year competition?

Kieran M. R. Hunt^{1,2} 

¹Department of Meteorology, University of Reading, UK

²National Centre for Atmospheric Sciences, University of Reading, UK

Introduction

Since the 1950s, artificial intelligence (AI), in various guises and levels of abstraction, has been used to create visual art. Early efforts typically comprised abstract computer graphics – such as a 1967 untitled piece by Frieder Nake¹ that now hangs in the Tate Modern – or required artistic styles to be hardcoded into the system – such as the AARON software written by Harold Cohen in the 1970s. However, in each case, the creativity came from the human rather than the computer.

This requirement for a high level of human interference changed with the development of generative adversarial networks (GANs; Goodfellow *et al.*, 2014). GANs are a type of artificial neural network (ANN)² used to generate new data samples from a given input dataset. GANs are composed of two sub-networks: a generator network that creates new data samples, and a discriminator network that tries to classify the generated samples as either real or fake. The two networks are trained together in a competitive manner, such that the generator network learns to create realistic data samples that fool the discriminator network, while the discriminator network becomes better at identifying fake data samples. GANs have seen considerable success in art and image generation, as well as special cases of photorealism (e.g. human faces³), but typically perform poorly in general cases of photorealism (e.g. Wang *et al.*, 2022).

¹<https://www.tate.org.uk/art/artworks/nake-no-title-p80817>

²ANNs are computational models inspired by biological neural networks, which use interconnected nodes and weights to process input data and map it to a desired output. By adjusting these weights according to identified errors, an ANN can learn to recognise patterns and produce accurate predictions.

³<https://www.thispersondoesnotexist.com/>



Figure 1. Comparison of two artificial intelligence artists, one based on a generative adversarial network (VQGAN + CLIP, top), and one based on a diffusion model (Stable Diffusion, bottom). Each was given the prompt: 'Award winning photograph of a foggy day in London'. Output is shown after 10, 25, 50 and 150 iterations of each model.

They have recently been surpassed in this regard by denoising diffusion probabilistic models (hereafter, diffusion models; Ho *et al.*, 2020). Diffusion models are essentially autoencoders⁴ trained to remove Gaussian noise from images; once trained, the denoising networks are sufficiently powerful that an image can be generated from pure noise. Both GANs and diffusion models can be trained to accept text prompts that 'guide' the generation or denoising towards a desired destination in some language embedding space.⁵ In each case, during model training, a caption describing the training image is compressed into this embedding using an additional pre-trained language model and fed in as a secondary input along with the image. Figure 1 compares attempted photorealistic image generation between a leading GAN and a leading diffusion model,

⁴Autoencoders are a type of neural network that are used to learn how to compress data. They do this by learning to encode the data into a smaller representation, and then decode it back into the original form.

⁵Language embedding spaces are mathematical spaces where words or phrases from a language are mapped according to some similarity metric.

given the same text prompt. Although the GAN produces an image that accurately fits the description of 'a foggy day in London', it is highly unlikely to fool any human into thinking it is a real photograph, even less so that it is 'award-winning'. In contrast, the diffusion model produces an image that might reasonably convince a human that it is a real photograph, even accurately rendering Big Ben and the Palace of Westminster.

In this article, we will explore the ability of three publicly available diffusion models to produce photograph-like images of weather themes capable of winning the Weather Photographer of the Year competition. We will compare the models against previous winners and discuss the relative abilities of each model across different themes before finishing with a Turing test – where the reader is invited to try to tell apart the real photographs from those created by the diffusion model 'AI artists'.

The diffusion models

This section provides a brief description of the three diffusion models used in this article. These three models were chosen because they were publicly available – ruling out, for example, Google's *Imagen* – and

consistently produced good quality output – beating, for example, crayon (also known as DALLE-mini) and LMU Munich’s Latent Diffusion.

DALLE-2 was developed by OpenAI, who also produced the open access language-embedding model CLIP (used by both Midjourney and Stable Diffusion, as well as VQGAN in Figure 1). Despite this, DALLE-2 uses the more recently developed GPT-3 as its language-embedding model. Although both DALLE-2 and GPT-3 are publicly available,⁶ neither are open access. DALLE-2 uses 3.5 billion parameters, a significant reduction from the 12 billion used in its predecessor; however, GPT-3 uses 175 billion parameters.

Midjourney was founded by David Holz. It is not open access but is publicly accessible through a dedicated Discord server.⁷ It uses CLIP for language embedding, with the underlying diffusion model constantly being refined. It is at a slight disadvantage to the other two diffusion models used, as Holz says: ‘we wanted it to kind of look beautiful, and beautiful doesn’t necessarily mean realistic If anything, actually we do bias it a little bit away from photos.’⁸

Stable Diffusion was developed by Stability.AI and made fully open access on 22 August 2022⁹ whereafter users could download and run it on their own computers. Prior to this, like Midjourney, it was publicly accessible through a Discord server, where the user could supply prompts to a bot. The most recent version, v1.4, uses 890 million parameters.

There are three final caveats before we start. First, for fair comparison throughout, all images are rendered square – either directly, or through cropping. When cropping actual photographs, I have tried to do so in as fair a manner as possible, preserving as much of the original content and artistic direction as I could. Second, I have indulged in minor ‘prompt engineering’ throughout. In other words, adding suggestions like ‘award-winning’ or the names of stock photo companies to encourage the AI artists to produce more professional-looking images. In most cases, without these additions, the AI would still produce a photorealistic depiction of the weather in the prompt, but they tended to be far less visually appealing. Third, in most cases I have had each artist produce several images for each prompt (never exceeding four) using a different random seed to generate the initial noise. This then requires one of several to be selected for use in the paper – typically I have selected those that I think best fit the competitive photography aesthetic, while

⁶<https://labs.openai.com/>

⁷<https://discord.com/channels/662267976984297473/>

⁸https://www.theregister.com/2022/08/01/david_holz_midjourney/

⁹<https://huggingface.co/CompVis/stable-diffusion>



Figure 2. Comparison of 2021 winning photographs with those produced by the three artificial intelligence artists. In each case, the text used to describe the photograph and its merits on the competition website was used as a prompt, with occasional modification. Overall winner prompt: ‘This photo can only be taken from one point. There is a small church on top of a hill in the town of Airuno, in the province of Lecco in Italy. Under the mist passes the River Adda. In the autumn months, on some days, it is possible to see this show with the first lights of sunrise’. Young photographer winner prompt: ‘Photo of beautiful clouds coming in right before a Kansas storm. Anyone who has experienced a severe thunderstorm knows about the eerie deep green/blue colour sometimes present as the storms approach’. Public choice winner prompt: ‘Lightning from an isolated storm over Cannes Bay. The judges commented that few storms are as beautiful as those isolated over water. The photographer was a perfect distance away from this storm to capture three things crucial for a winning photo composition: the sky, the storm, and the water’.

rejecting the handful that contain obvious giveaway artefacts or errors.

Comparison with previous winners

For the first task (Figure 2), we compare our three AI artists to three winners of the 2021 Weather Photographer of the Year competition – the overall winner and the winners of the youth photographer and public choice categories. In each case, the AI is fed only the description of the winning photo given on the RMetS website¹⁰, modified slightly in the young photographer case to include ‘photo of’ at the beginning.

The overall winner was a photo of a foggy autumn morning in northern Italy taken from a hilltop church by Giulio Montini. The prompt mentions only that a church is involved rather than explicitly defining it as the viewpoint, and that ambiguity leads Midjourney and Stable Diffusion to produce photographs of an Italian-style rural church surrounded by morning mist. DALLE-2 produces an image much closer to the winner, capturing a foggy

valley (albeit containing a church tower) lit by a low elevation morning sun.

The young photographer winner was a photo of a supercell building over a farm in Kansas taken by Phoenix Blue. The prompt does not mention a farm, but does include Kansas, and the two AI artists that include some landscape (Midjourney and Stable Diffusion) produce scenery identifiable as the Great Plains. All three AI artists capture the greenish hue associated with developing severe convection, but Stable Diffusion is probably closest to the winner in terms of producing an authentic supercell cloud structure. Note that the prompt does not explicitly mention a supercell, so the AI artists have inferred this structure based on it being a severe thunderstorm over Kansas.

The public choice winner (and overall runner-up) was a photo of a lightning strike off the southeast coast of France by Serge Zaka. The prompt is clear and detailed, and all three AI artists capture the scene correctly with DALLE-2 and Stable Diffusion producing realistic looking clouds, water and lightning. Midjourney captures realistic looking clouds and sea, but the lightning looks more like a firestorm from a fantasy film – though it is at least accurately reflected in the ocean.

¹⁰<https://www.rmets.org/news/winners-announced-weather-photographer-year-2021>

These results highlight the importance of using the right prompt and right theme with the right AI artist, and show that this type of direct comparison is probably not a fair competition.

Exploring prompts and themes

So, if we are going to get these AI artists to produce competition-worthy photographs, we need to get a feel for the themes and prompts that suit their particular style. After all, one would not ask Annie Leibovitz to take landscape photos, nor Ansel Adams to take portraits. In Figure 3, we give each of the three AI artists a more ‘artist-friendly’ prompt that describes the composition and style in more detail. They are given one prompt for each of four themes (extreme weather, landscape, human interest and macro/close-up) that broadly capture all the Photographer of the Year competition finalists over the few years.

We can see that each artist has a different strength. DALLE-2 is excellent at capturing physics and structural features, highlighted in its render of the breaking wave hitting a lighthouse. This is challenge because not only are breaking waves turbulent and extremely complex, but the artist must also correctly illuminate the structure – containing thousands of translucent or diffractive particles – for it to be believable. Not only that, but DALLE-2 simulates an interaction between the two, showing the wave coming around each side of the lighthouse.

Midjourney is the most creative with the prompts and has perhaps the best lighting of the three AI artists. For prompt 2, the bridge is in the distance with the frosty field in the foreground beautifully (and correctly, given the roughness of the surface) illuminated by the rising sun’s yellow light. The wave is also lit in a visually striking manner, and, unlike the other two artists, the Indian girl is rendered as a silhouette with the monsoon background as a focus – an inventive way to respond to the request for ‘candid’ framing.

Stable Diffusion excels at composition. Regardless of the ultimate render quality, each of the four prompts was framed in the manner most consistent with a human photographer. This is particularly highlighted both by the clever choice of depth-of-field in the plant bud shot; the still, reflective stream passing under the frosty bridge; and the portrait-like framing of the Indian girl, including a realistic render of her wet clothing. All three artists can produce photorealistic renders of the given prompts, but it is clear that to fool the human eye – or even win a competition – we must choose the right combination of artist, theme and prompt.



Figure 3. Comparison of the three artificial intelligence artists across four different weather photography themes (extreme, landscape, human interest, and macro). Prompt 1: ‘Award-winning photograph of large waves crashing into a lighthouse in Devon during the peak winds of Storm Eunice’. Prompt 2: ‘Award-winning photograph of a small stone bridge in the English countryside, covered in frost, lit by an early sunrise. The photographer got up early and waited for the clouds to clear before capturing this moment’. Prompt 3: ‘Award-winning candid black and white photo of a young Indian woman in Varanasi smiling and dancing in the first monsoon rains’. Prompt 4: ‘Close up photograph of a single bud emerging from dried, cracked ground with some dead grass after a long drought, bokeh trees in the distance, hot summer day, Photofest, trending, 4K’.

The final test

In this final section, you – the reader – are invited to judge. Sixteen images are given in Figure 4, five of which are real photographs that were shortlisted finalists in the 2021 Photographer of the Year competition. The other 11 have been generated by the three AI artists (four by DALLE-2, two by Midjourney, five by Stable Diffusion). When you have made your decision, you can check your answer by looking at the final word in the second paragraph of the conclusion, which comprises the five letters labelling the real photographs. The text prompts and models used for each

photograph are given in the Supporting Information Notes S1.

Concluding remarks

In this short article, we explored the ability of three publicly available diffusion models (DALLE-2, Midjourney and Stable Diffusion) at creating realistic weather-themed photographs that could compete in a photography competition. We discussed the relative strengths and weaknesses of each of the three AI artists before finishing with a Turing test, where the reader was asked to choose which 5 of 16 photographs were shortlisted finalists in the 2021 Weather Photographer



Figure 4. Sixteen ‘photographs’ of weather. Five are real photographs – shortlisted finalists in the Weather Photographer of the Year 2021 competition. The other 11 have been created by the three artificial intelligence artists discussed in this paper. As a Turing test, the reader is invited to guess which five are the real photographs.

of the Year competition, and which 11 were created by the diffusion models.

It is, of course, not that simple. There is still a reasonable degree of human creativity required – in dreaming up the text prompts, in refining them to make the generated images more suitable, and in selecting the best images from those produced to take forward. Nor are these artists perfect – many generated images still contain artefacts that immediately give them away as fakes – and they are still far better at creating artistic stylisations than photorealistic images. But, these gaps are reducing with every new hardware and software development. Eventually, when it comes to telling the dif-

ference between human and AI-generated photographs, we will be blind.

GANs and diffusion models are not limited to producing images and have many potential applications in other fields, including meteorology and climate science. Such applications include downscaling climate model output (Cheng *et al.*, 2021), ensemble weather prediction (Bihlo, 2021), nowcasting (Rüttgers *et al.*, 2022) and even storm surge models (Lütjens *et al.*, 2020). Generative models can also now produce convincing text output. To demonstrate this, we used GPT-3¹¹ to produce the first

¹¹<https://beta.openai.com/playground>

paragraph of the introduction as well as the definitions used in footnotes 2, 4 and 5.

Acknowledgement

Kieran M. R. Hunt is supported by a NERC Independent Research Fellowship (MITRE; NE/W007924/1). Aside from the figures shown, no new data were created for this article.

References

- Bihlo A.** 2021. A generative adversarial network approach to (ensemble) weather prediction. *Neural Net.* **139**: 1–16.
- Cheng J, Liu J, Kuang Q et al.** 2021. DeepDT: generative adversarial network for high-resolution climate prediction. *IEEE Geosci. Remote Sens. Lett.* **19**: 1–5.
- Goodfellow I, Pouget-Abadie J, Mirza M et al.** 2014. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**: 1–9.
- Ho J, Jain A, Abbeel P.** 2020. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **33**: 6840–6851.
- Lütjens B, Leshchinskiy B, Requena-Mesa C et al.** 2020. Physics-informed gans for coastal flood visualization. *arXiv:2010.08103*.
- Rüttgers M, Jeon S, Lee S et al.** 2022. Prediction of typhoon track and intensity using a generative adversarial network with observational and meteorological data. *IEEE Access* **10**: 48434–48446.
- Wang Z, Liu W, He Q et al.** 2022. CLIP-GEN: language-free training of a text-to-image generator with CLIP. *arXiv:2203.00386*.

Correspondence to: K. M. R. Hunt

k.m.r.hunt@reading.ac.uk

© 2022 The Authors. *Weather* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

doi: 10.1002/wea.4348

Supporting Information

Notes S1. Text prompts and models used for Figure 4.