Fig. 6. Example of applying a "smile vector" with an ALI model [19]. On the left hand side is an example of a woman without a smile and on the right a woman with a smile. A $\mathbf{z}$ value for the image of the woman on the left is inferred, $\mathbf{z}_1$ and for the right, $\mathbf{z}_2$. Interpolating along a vector that connects $\mathbf{z}_1$ and $\mathbf{z}_2$, gives $\mathbf{z}$ values that may be passed through a generator to synthesize novel samples. Note the implication: a displacement vector in latent space traverses smile "intensity" in image space.

useful and is now a common practice to improve image quality. This idea of GAN conditioning was later extended to incorporate natural language. For example, Reed et al. [43] used a GAN architecture to synthesize images from text descriptions, which one might describe as *reverse captioning*. For example, given a text caption of a bird such as "white with some black on its head and wings and a long orange beak", the trained GAN can generate several plausible images that match the description.

In addition to conditioning on text descriptions, the Generative Adversarial What-Where Network (GAWWN) conditions on image location [44]. The GAWWN system supported an interactive interface in which large images could be built up incrementally with textual descriptions of parts and user-supplied bounding boxes (Fig. 7).

Conditional GANs not only allow us to synthesize novel samples with specific attributes, they also allow us to develop tools for intuitively editing images – for example editing the hair style of a person in an image, making them wear glasses or making them look younger [35]. Additional applications of GANs to image editing include work by Zhu and Brock et al. [2], [45].

### C. Image-to-image translation

Conditional adversarial networks are well suited for translating an input image into an output image, which is a recurring theme in computer graphics, image processing, and computer vision. The `pix2pix` model offers a general purpose solution to this family of problems [46]. In addition to learning the mapping from input image to output image, the `pix2pix` model also constructs a loss function to train this mapping. This model has demonstrated effective results for different problems of computer vision which had previously required separate machinery, including semantic segmentation, generating maps from aerial photos, and colorization of black and white images. Wang et al. present a similar idea, using GANs to first synthesize surface-normal maps (similar to depth maps) and then map these images to natural scenes.

CycleGAN [4] extends this work by introducing a cycle consistency loss that attempts to preserve the original image after a cycle of translation and reverse translation. In this formulation, matching pairs of images are no longer needed for training. This makes data preparation much simpler, and opens the technique to a larger family of applications. For example, artistic style transfer [47] renders natural images in the style of artists, such as Picasso or Monet, by simply being trained on an unpaired collection of paintings and natural images (Fig. 8).

### D. Super-resolution

Super-resolution allows a high-resolution image to be generated from a lower resolution image, with the trained model inferring photo-realistic details while up-sampling. The SRGAN model [36] extends earlier efforts by adding an adversarial loss component which constrains images to reside on the manifold of natural images.

The SRGAN generator is conditioned on a low resolution image, and infers photo-realistic natural images with 4x up-scaling factors. Unlike most GAN applications, the adversarial loss is one component of a larger loss function, which also includes perceptual loss from a pretrained classifier, and a regularization loss that encourages spatially coherent images. In this context, the adversarial loss constrains the overall solution to the manifold of natural images, producing perceptually more convincing solutions.

Customizing deep learning applications can often be hampered by the availability of relevant curated training datasets. However, SRGAN is straightforward to customize to specific domains, as new training image pairs can easily be constructed by down-sampling a corpus of high-resolution images. This is an important consideration in practice, since the inferred photo-realistic details that the GAN generates will vary depending on the domain of images used in the training set.

## VII. Discussion

### A. Open Questions

GANs have attracted considerable attention due to their ability to leverage vast amounts of unlabelled data. While