Recovering the 3D shape of an object from 2D images

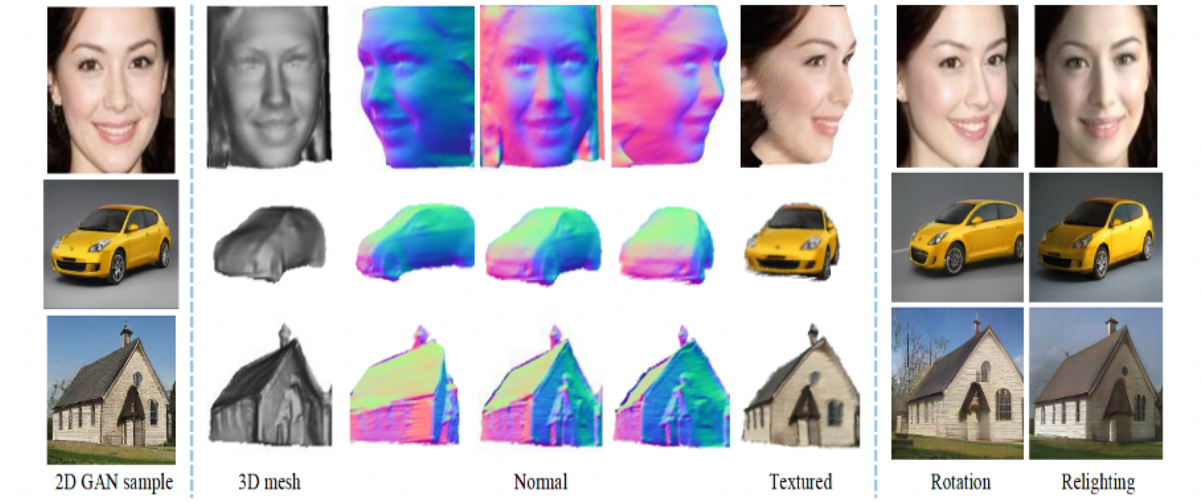

Figure 1 Images with different effects

Natural images are projection of 3D object on a 2D image plane. To generate the 3D object model from the 2D images with different viewpoints and lighting as shown in the above figure 1. The knowledge gained from the 2D images is used to recover the 3D shape of the object. The methodologies of recovering 3D shape varies with the source of 2D images.

Existing Methods of GANs

Generative adversarial networks can model the 2D natural image manifold of diverse object categories with high fidelity. GAN's can shift object in generated images which can change the viewpoint of object in GAN image manifold. This phenomenon motivates the possibility of reconstructing the 3D shape from single 2D image. However, due to either heavy memory consumption or increased difficulty in training, there are still gaps between their image qualities and those of 2D GANs. So, there is need for the memory efficient extraction of knowledge from the 3D shape and texture generation with high quality. There has been works trying to discover the latent space directions that manipulate the image content in a 3D-controllable manner using 3D aware generative manipulation models. The lighting factors of the images are ignored with borrowing the external 3D models. And it cannot generalize beyond the faces, cars etc. In unsupervised learning of 3D shapes, it doesn't need manual annotations and lacks multiple viewpoints and lightning instances. To tackle it they adopted the shapes from the external 3D shapes or 2D images which learns viewpoint and shape via autoencoders that explicitly model the rendering process.

Methodologies for Recovering 3D Shape

The methodology of using GAN to learn 3D shapes from images which rely on explicitly modelling 3D representation and rendering during training. Due to either heavy memory consumption or additional training difficulty brought by the rendering process, the qualities of their generated samples notably lag 2D GAN counterparts. Another methodology of learning 3D shapes from the images learns the viewpoint and shapes of images in 'analysis by synthesis' manner. Although this methodology has given great results, it assumes all objects shapes as symmetric which doesn't actual work for many objects.

Proposed model

Although the knowledge from the 2D images with different instances of viewpoint and lighting, the challenge is to discover well disentangled semantic directions from the image manifold. Manually inspecting and annotating the images manifold that control viewpoint and lighting is laborious and time consuming. To tackle this the idea of using ellipsoid for generating different images. On the image generated by GAN an ellipsoid is employed and as its first 3D object shape and render several unnatural images called pseudo samples. Pseudo samples are randomly sampled viewpoints and lighting conditions. By reconstructing them using the GAN, these pseudo samples could guide the original image towards the sampled viewpoints and lighting conditions in the GAN manifold, producing a number of natural-looking images, called projected samples. The framework of the creating pseudo samples and projected samples is shown as below in figure 2.

The following are the steps for reconstructing the 3D shape from the 2D image using the GAN2shape with which the inversion strategy works for not only GAN images but also for real natural images.

1) We present the first attempt to reconstruct the 3D object shapes using GANs that are pre-trained on 2D images only. Our work shows that 2D GANs inherently capture rich 3D knowledge for different object categories and provides a new perspective for 3D shape generation.

2) Our work also provides an alternative unsupervised 3D shape learning method and does not rely on the symmetry assumption of object shapes.

3) We achieve highly photo-realistic 3D-aware image manipulations including rotation and relighting without using any external 3D models.

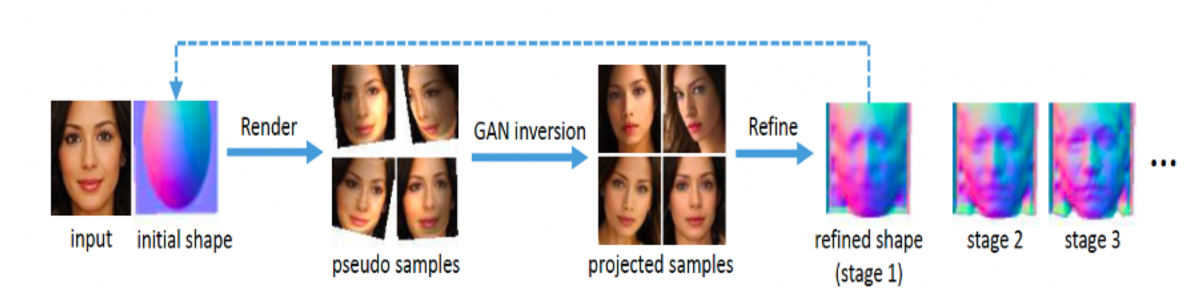

Figure 2 Outline of the framework

Before getting into usage of GAN 3D shape recover the details GAN image generation is discussed. GAN consists of generator which maps a latent vector z to an image and a discriminator D that distinguishes between generated images and real ones. GANs trained on natural images can continuously sample realistic images, thus could be used as an approximation of the natural image manifold. However, the GAN2style consists of two parts, a mapping network that maps the latent vector in the input latent space to an intermediate latent vector, and a synthesis network that maps to the output image.

There are steps for the 2D GAN image to learn the 3D shape of the object as shown below in the Figure 3.

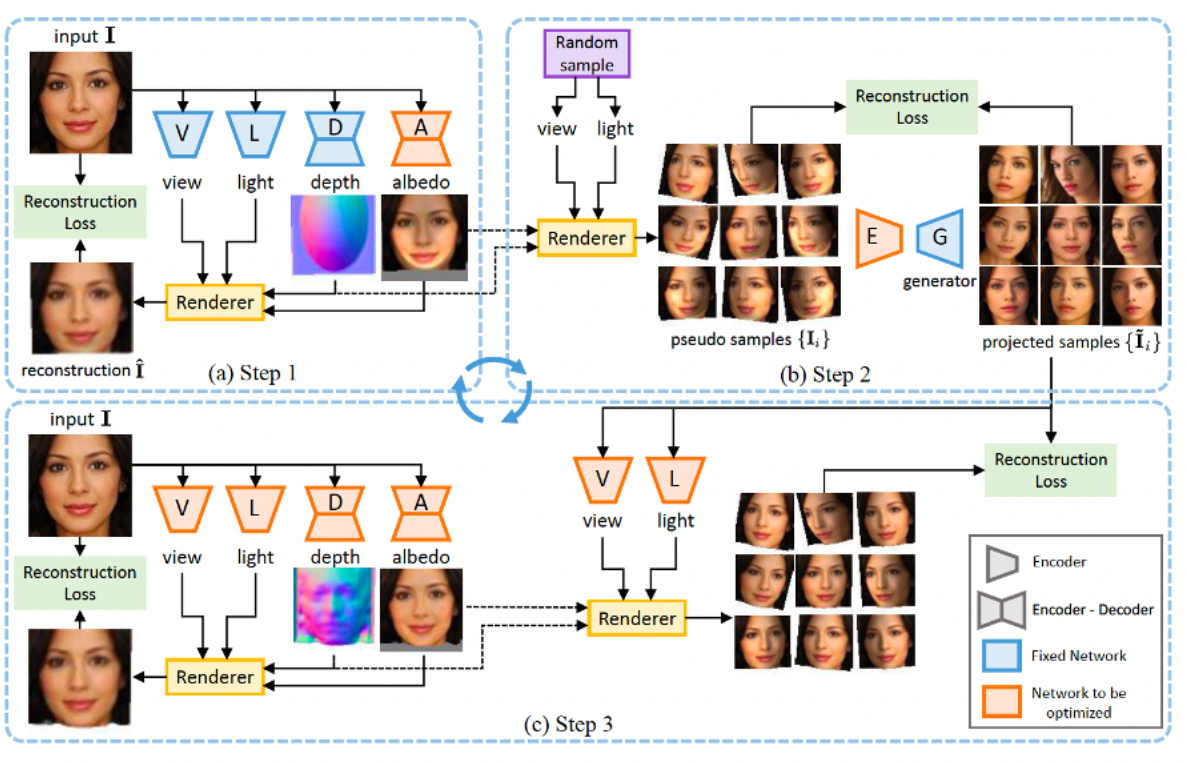

Figure 3 steps of 3D shape recovering from the 2D GAN image

Step 1: Using a Weak Shape Prior

With single image instance to explore images with many viewpoints and lighting to the GAN image manifold could resort to ellipsoid as weak prior to create pseudo samples. For the image sample I from the GAN, we initialize the depth map d to have an ellipsoid shape. With an off-the-shelf scene parsing model we could position the ellipsoid to be roughly aligned with the object in the image. Then the albedo network A is optimized with the reconstruction objective L(I, ˆI), where ˆI is calculated and L is a weighted combination of L1 loss and perceptual loss. The viewpoint v and lighting l are initialized with a canonical setting, *i.e.*, v0 = 0 and lighting is from the front. This allows us to have an initial guess about the four factors d0(depth), a0(albedo), v0(viewpoint),l0(lighting), with d0 being the weak shape prior.

Step 2: Sampling and Projecting to the GAN Image Manifold.

With the initialisation from the first step the pseudo samples with different lighting directions and viewpoints are created. Although these pseudo samples have unnatural distortions and shadows, they provide cues on how the face rotates and how the light changes. In order to leverage such cues to guide novel viewpoint and lighting direction exploration in the GAN image manifold, we perform GAN inversion to these pseudo samples, *i.e.*, reconstruct them with the GAN generator . Specifically, we train an encoder that learns to predict the intermediate latent vector for each sample. Generator could regularize the projected samples to lie in the natural image manifold, thus fixing the unnatural distortions and shadows in the pseudo samples.

Step 3: Learning the 3D Shape

As all the images have same content with different viewpoints and lighting directions makes it possible to learn the underlying 3D shape with aforementioned photo-geometric autoencoding model. The viewpoint network, lighting network , the depth network and albedo network are all together optimised to learn the 3D shape. The reconstruction loss is minimised on when all the predictions are inferred. The unnatural distortions and shadows that exist in the pseudo samples are mitigated in the reconstructions.

Conclusion

We have presented the first method that directly leverages off-the-shelf 2D GANs to recover 3D object shapes from images. We found that existing 2D GANs inherently capture sufficient knowledge to recover 3D shapes for many object categories, including human faces, cats, cars, and buildings. Based on a weak convex shape prior, our method could explore the viewpoint and lighting variations in the GAN image manifold, and exploit these variations to refine the underlying object shape in an iterative manner.

References

https://arxiv.org/pdf/2011.00844v4.pdf