The main aim of this study is to employ an advanced technique known as Pix2Pix Generative Adversarial Networks (GANs) to create high-quality unmasked images. These images are intended to reveal facial features that are typically hidden by masks.

To put it differently, the problem we are addressing can be framed as the removal of mask obstructions from images. Prior methods, such as "Region filling and object removal by exemplar-based image inpainting," "Image melding: combining inconsistent images using patch-based synthesis," and "Scene completion using millions of photographs," removed masked objects and filled the gaps with similar patches or patterns, but they had limitations in producing realistic results.

Generative Adversarial Networks (GANs), introduced by Ian J. Goodfellow and his colleagues, involve a training process where a generator creates data from noise, and a discriminator distinguishes between real and fake data. This feedback loop helps the generator improve its output.

Conditional GANs (cGANs), use specific information to learn the relationship between related images. Pix2Pix GAN is an example of a conditional GAN that depends on an input image to generate an output. This is unlike traditional GANs that generate from noise. Conditional GANs have been used in various applications, including generating cartoons, adding shadows to virtual objects, predicting breast cancer prognosis, optimizing autonomous robot paths, and more.

The primary application of Pix2Pix GAN is in Image-to-Image translation, using a U-Net generator model. Pix2Pix GAN is chosen for this problem because it excels in the following ways:

1. Conditional Input: It uses a source image as a condition, enabling precise image-to-image translations tailored to the task.
2. Loss Functions: Pix2Pix GAN combines adversarial loss and pixel-level similarity loss, promoting realism and accuracy in translations.

The generator model's primary function is to produce unmasked images of individuals by utilizing a U-Net architecture. This architecture consists of two key components: the encoder and the decoder. The interconnections between the layers in these components, referred to as skip connections, enable the seamless flow of information. The encoder is responsible for progressively reducing the image size while capturing and extracting finer details from the input image. In contrast, the decoder performs upsampling to reconstruct the image by utilizing the extracted details and incorporating supplementary data transmitted through the skip connections.

In contrast, the discriminator model has the responsibility of differentiating between real, unaltered images and the artificially created ones produced by the generator. In this particular case, the discriminator employs a PatchGAN architecture, which means it assesses the genuineness of smaller sections within the image and computes the likelihood of the image being authentic or artificially generated. This likelihood value acts as a form of input to the generator model, motivating it to generate more realistic, unaltered images in order to outsmart the discriminator. This continuous exchange of information is commonly referred to as the Adversarial loss.

In addition to the Adversarial loss mentioned earlier, the Pix2Pix GAN training process includes an extra component known as the L1 loss. This loss function operates on a per-pixel basis, customized to the task's requirements, with the objective of reducing the differences between the artificially generated unmasked images and the authentic ones at a fundamental pixel level.

Throughout the training process, a common discriminator model assesses both the generated unmasked image paired with the input masked image and the authentic unmasked image paired with the input masked image. This configuration aids in the discriminator's ability to differentiate between genuine and artificially created results. The generator's loss function includes an adversarial component influenced by the discriminator's judgments and a task-specific component, such as the L1 loss, which contributes to the creation of more lifelike images. The gradients derived from these loss functions guide the optimization of variables for both the generator and discriminator. This iterative process enhances the discriminator's capacity to distinguish real images from synthetic ones and empowers the generator to produce increasingly genuine visual outputs.

The Pix2Pix GAN model's unmasked image outputs demonstrate its ability to remove obscured mask objects and fill vacant areas with appropriate low-level and high-level details. To validate our findings, we conducted a survey and collected preference scores from a group of participants to assess the accuracy of the model's unmasked predictions. We opted for this approach because conventional metrics like per-pixel mean-squared error fail to adequately capture the qualities of the final image.

In this evaluation, 46 participants took part, and our model received a likability index score of 95.5, indicating that volunteers perceived its generated images as realistic. This subjective assessment offers a precise and accurate measure of synthesized image quality, capturing the subjective experience of the human visual system, which is often overlooked by standard metrics.

Previously, we introduced a Pix2Pix model tailored to produce unmasked images from corresponding masked photographs using conditional adversarial networks and the U-net architecture. The results we obtained confirm that our model effectively generates unmasked images, and the subjective evaluation conducted with diverse volunteers reaffirms its high likability. While our model performs admirably with the masked objects within the dataset, it encounters challenges when tasked with generating high-quality unmasked images from input photos with different types of masks. Going forward, our primary focus will be on enhancing the model's adaptability to accommodate masks of varying shapes and colors, thereby making it more versatile.