**Model Report (Handover Document)**

**1. ESRGAN (Enhanced Super-Resolution GAN)**

**Purpose:**  
ESRGAN was initially chosen for its ability to generate high-quality super-resolution images, particularly for upscaling low-resolution images by a factor of 4.

**Limitations/Challenges:**

* **Over-focus on texture generation:** The ESRGAN model tends to generate visually appealing textures but can lose finer details. In cases where accurate restoration of structural details (e.g., lines, edges) is more critical than perceptual quality, ESRGAN underperformed.
* **Poor handling of real-world noise:** ESRGAN is typically trained on clean datasets, and in this project, real-world noisy images caused artifacts and distortion in the super-resolved images.
* **Mismatch with the task goal:** ESRGAN’s main purpose is super-resolution, and it struggled with the broader restoration needs of this task, which involved more than just enhancing resolution (e.g., deblurring or removing noise).

**2. UNet**

**Purpose:**  
UNet, with its encoder-decoder structure, was employed for image restoration because of its strengths in tasks like segmentation and general-purpose image enhancement.

**Limitations/Challenges:**

* **Over-segmentation bias:** UNet’s architecture, being designed originally for segmentation, tended to split images into regions and failed to fully capture global context, which is essential for restoring fine details across the entire image.
* **Insufficient detail recovery:** In this task, the level of detail in some complex textures or patterns could not be restored by the UNet model, leading to blurry or incomplete restorations.
* **Handling high-frequency noise:** UNet struggled with high-frequency noise and intricate image artifacts. The simple convolutions were insufficient for dealing with the complexities in the dataset, leading to unsatisfactory denoising performance.

**3. DeepDeblur**

**Purpose:**  
DeepDeblur is specifically designed for motion deblurring, which made it a good candidate for the task of removing blur from images.

**Limitations/Challenges:**

* **Ineffective with non-uniform blur:** The model is tailored towards images with uniform motion blur. In this project, the blur in the images was often non-uniform and varied in intensity, leading to suboptimal results. The model couldn’t generalize well to handle various types of blur (e.g., Gaussian, radial).
* **Complexity of image content:** The diverse and intricate nature of the content within the images posed challenges for DeepDeblur, which led to a failure in recovering sharp edges in certain regions.
* **Long training times with diminishing returns:** Despite experimenting with extended training sessions and hyperparameter tuning, performance gains became minimal after a point, without achieving the desired deblurring effects.

**4. DeblurGAN**

**Purpose:**  
DeblurGAN uses a GAN-based approach to deblur images, which was expected to provide better results by leveraging the discriminator network for enhanced realism in the deblurred images.

**Limitations/Challenges:**

* **GAN instability:** Training GANs, including DeblurGAN, can be unstable and prone to mode collapse. The generated results fluctuated in quality, and in several cases, the model failed to converge to an optimal solution during training.
* **Artifact generation:** While DeblurGAN was able to remove some level of blur, it introduced undesirable artifacts in the process. These artifacts appeared in areas of the image that should have remained unchanged, resulting in distortions.
* **Inconsistent performance across different image types:** The model performed well on simpler images but failed to generalize across a wider variety of textures and noise levels present in the dataset.

**5. TextZoom**

**Purpose:** TextZoom was explored for text image super-resolution, aiming to upscale low-resolution text images for improved readability**.**

**Limitations/Challenges:**

* **Complex Dependencies and Configuration:**The implementation of TextZoom required multiple dependencies and complicated configuration setups, making it challenging to integrate into the existing workflow. Additionally, configuring the model across different hardware setups caused inconsistencies in the results.
* **Training Requirements:**TextZoom required extensive fine-tuning and training datasets for optimal performance, which was beyond the available resources for this project. Given these difficulties, the decision was made to discard TextZoom for this task.

**5. Image Preprocessing (Python ImageEnhancement Script)**

**Purpose:** The image preprocessing method aimed to enhance the contrast between characters and the background in order to improve visual clarity. This preprocessing is especially useful for low-quality images where characters are faint or the background is noisy.

**Methodology:**

1. **Image Grayscale Conversion:**Each input image is first converted into a grayscale format using the Pillow library, ensuring a more straightforward manipulation of brightness and contrast levels.
2. **Contrast Enhancement:**The contrast of the grayscale image is then increased by a factor of 3.0, darkening the characters while brightening the background. This technique improves visibility, making the text stand out from the background**.**
3. **Sharpness Adjustment:**The sharpness of the images is also enhanced (with a factor of 2.5), which helps to define the edges of characters and improve the clarity of text in the images.

**Results:** The preprocessing script produced consistent results across a variety of images. The enhanced images showed significant improvement in text readability and edge sharpness.

**Limitations:** Although this method worked well for most images, its performance may degrade in cases where images have extreme blur or very low resolution, as further deblurring or upscaling might be necessary.

**Conclusion:**

While each of the models attempted had its strengths, none of them fully addressed the challenges presented by this particular image restoration task. The complexities of the dataset, including noise, blur variation, and detailed textures, required a more tailored approach that these models could not provide out of the box.

Future work will involve exploring custom model architectures or advanced fine-tuning techniques, such as:

* **Hybrid models** combining the strengths of multiple networks (e.g., combining deblurring and denoising).
* **Transformers-based architectures** that may be better suited to capture the long-range dependencies and patterns in the images.

**Next Steps:**

* **Use of Transformer-Based Models:**  
  Transformer-based architectures like **SwinIR** or **ViT** could potentially perform better for text enhancement tasks. These models can capture long-range dependencies and are less likely to suffer from localized focus issues, which might help in handling varying text styles and orientations.
* **Hybrid Models for Blur Removal and Denoising:**  
  A combination of deblurring and denoising models in a hybrid architecture could enhance the results further. For instance, using **DeblurGAN** in conjunction with a denoising autoencoder might result in clearer images without introducing artifacts.
* **Explore Lightweight Super-Resolution Models:**  
  Given the failure of TextZoom, a lightweight super-resolution model such as **EDSR** (Enhanced Deep Super-Resolution) could be experimented with. This model has fewer dependencies and configurations compared to TextZoom and might yield better results for real-time or large-scale tasks.