1. In 400 words or less, describe the problem the authors are trying to solve, the solution(s) proposed by the author, and their main considerations.

The authors are primarily trying to solve the problem of enhancing low-light images without creating unrealistic artifacts. The solution proposed is their novel method called Zero-Reference Deep Curve Estimation (Zero-DCE), which frames the problem as an image-specific curve estimation using deep neural networks.

The solution is considered novel because they use “Zero-Reference” to train the model. Specifically, neither paired or unpaired references images used in previous CNN-based or GAN-based methods were used. Paired reference images allow the model to take in raw input images and learn from the same output image but enhanced optimally to provide a reference to learn from. This is often used in supervised CNN-based methods. Unpaired reference images do not have this input and output link, in other words the model simply learns from a set of carefully selected “good” images as reference and this is often used in GAN-based methods.

Zero-DCE however does not require any sort of reference for training but instead utilizes a set of 4 custom loss functions: Spatial Consistency Loss, Exposure Control Loss, Colour Constancy Loss, and Illumination Smoothness Loss. These loss functions have to be differentiable to be used by backpropagation during model training and they all try to optimize for a certain aspect of the low-exposure images. They all eventually help provide an optimal image-specific enhancement curve, closely related to how photo or video editors enhance the images and videos using the “tone curve”.

The 2 main considerations for the authors when it comes to trying to improve from previous methods are firstly of course, the “performance”, how well the low-exposure images are enhanced for better qualitative and quantitative metrics, which can mean a natural-looking image without unrealistic artifacts such as under- or over-exposure or colour deviation. Secondly, the other main consideration is the efficiency of the model training and inference processes. Since reference images are needed for previous techniques, which is resource intensive and might not be realistic since they might be artificially generated, having a fast model that is also easily trained and have reduced resource requirements will make their solution a stark improvement over other methods.

2. The paper claims to perform better than other methods. What quality assessment metric were used in this paper? Describe in 400 words or less the quality metric used for evaluation and how they were computed.

Apart from the 4 custom loss functions used during model training, the authors used both qualitative and quantitative metrics to compare against previous methods to conclude that their method is more performant.

They employed a User Study to quantify the subjective visual quality from 15 persons to independently score the outputs from 1 to 5 based on presence of artifacts or over-/underexposure, colour deviation, and unnatural texture and obvious noise. They also performed a non-reference Perceptual Index, usually used to measure other image restoration tasks such as image dehazing, to measure the perceptual quality (lower is better).

Next, they used 4 quantitative metrics, 3 for the image quality and 1 for the model inference runtime in seconds, is measured with input sizes of 1200x900x3 on Nvidia GTX 2080Ti GPUs. The 3 image quality metrics are Peak Signal-to-Noise Ratio (PSNR, dB), Structural Similarity (SSIM), and Mean Absolute Error (MAE).

PSNR is computed using the formula: $PSNR = 10 * log10(MAX^2 / MSE)$, where MAX is the maximum pixel value of the image, and MSE is the mean squared error between the reference and distorted images.

SSIM is an image quality assessment metric that measures the structural similarity between two images, ranging between -1 and 1, with 1 indicating perfect similarity. The SSIM values are computed using the formula:

$$SSIM(x,y) = (2 * \mu x * \mu y + c1) * (2 * \sigma xy + c2) / (\mu x^2 + \mu y^2 + c1) * (\sigma x^2 + \sigma y^2 + c2)$$

Where x and y are the reference and distorted images, μ for mean, σ for standard deviation, σxy for covariance, and c for constants to avoid division by zero.

MAE measures the average difference between the predicted and ground truth values. In this case, the predicted values are the enhanced brightness values obtained using the Zero-DCE method. The formula is as follows:

MAE = (1 / n) * sum(|Ie(i,j) - Ib(i,j)|), for i = 1 to m, j = 1 to n

Where I is the input image, and Ib and Ie are the brightness values before and after enhancement. The absolute difference |Ie(i,j) - Ib(i,j)| is taken for each pixel. The sum of differences is then divided by the total number of pixels to get the average difference.

All of these metrics were shown in the paper to be an improvement over previous methods, thus backing up their claim that their solution is “better”.

3. Describe in 400 words or less a possible use-case for HomeTeam departments that would require the technology presented in this paper.

Having served as a frontline GRF officer in SPF, I am fairly familiar with HomeTeam. Considering the primary focus to ensure safety and security in Singapore, many of HomeTeam's departments can benefit from this technology.

Specifically, since this paper proposes a training- and inference-efficient solution to enhance low-exposure images, which can then be used to further improve face detection models, as highlighted in the paper, Zero-DCE can be employed in surveillance systems, which deal with image and video data. Especially considering how safety and security is more critical yet harder to manage in low-light environments, this solution can improve the data received by surveillance cameras in these situations.

Moreover, since these systems often handle real-time video feed, having a fast and lightweight model that can be deployed and provide real-time inference will help achieve an efficient monitoring system, which the solution claims to be 3x faster than even the second-best existing method, EnlightenGAN.

Hence, with these features, this solution can be used in any systems that use surveillance cameras, such as prison cells, referencing Singapore Prison Service's "Prison Without Guards" initiative, and dimly lit or low footfall environments such as alleyways.

Furthermore, this low-exposure enhancement model can be part of the early stages of a complete surveillance pipeline starting with the image or video captured input, to image pre-processing and enhancement using this and other solutions, to downstream models such as face detection to detect humans and faces, pose estimation to detect or even predict human movements, and object detection or segmentation to locate and classify objects that might potentially be dangerous, and finally all these as part of an automated alert and observation system especially for low-light environments.

Such a pipeline consisting of multiple of such solutions including this novel one will undoubtedly optimize Singapore's surveillance and monitoring systems and be scaled up and deployed to multiple locations and use-cases, thereby strengthening and solidifying Singapore's image as one of the safest places in the world.