# Supplementary Material for
# LLVIP: A Visible-infrared Paired Dataset for Low-light Vision

Anonymous ICCV 2021 submission
Paper ID 5

## Contents

**1. Detailed Introduction of Fusion Metrics** 1

**2. More Examples of LLVIP and Their Fusion Results** 2

## 1. Detailed Introduction of Fusion Metrics

We objectively evaluate the performances of different fusion methods with some popular metrics in the paper. Here we introduce the metrics in more detail, including entropy (EN), mutual information (MI) [3, 4] series, structural similarity (SSIM) [5], $Q_{abf}$ [2] and visual information fidelity for fusion (VIFF) [1].

EN is defined based on information theory, which measures the amount of information the fused image contains. Mathematically, EN is defined as follows:

$$EN = -\sum_{l=0}^{255} p_l log_2 p_l \tag{1}$$

where $l$ is the gray value of pixels, $p_l$ is the normalized histogram of corresponding gray level in the fused image. The larger entropy is, the more information fused image contains, and the better performance fusion achieves.

MI [3] is the most commonly used objective metric for image fusion. Fusion factor(FF) and fusion symmetry(FS) [4] are concepts based on MI. FF is defined as:

$$FF = I_{VF} + I_{IF} \tag{2}$$

where $I_{VF}$ and $I_{IF}$ respectively represent MI between visible image V and fused image F, and between infrared image I and fused image F. A high FF value indicates that the fused image contains a considerable amount of information that exists in both images.

Our modified FS is defined as:

$$FS = \frac{I_{VF}}{I_{VF} + I_{IF}} \tag{3}$$

where FS represents the proportion of mutual information $I_{VF}$ in FF. If FS is greater than 0.5, the fused image contains more visible image information; if FS is less than 0.5, the fused image contains more infrared image information.

Normalize mutual information $Q_M I$ is defined as:

$$Q_{MI} = 2(\frac{I_{VF}}{H_V + H_F} + \frac{I_{IF}}{H_I + H_F}) \tag{4}$$

where $H_V$, $H_I$, $H_F$ are entropy of visible image, infrared image and fusion image. The greater the value of $Q_M I$, the more information is obtained from the source images, and the better the fusion effect is.

SSIM [5] is a perceptual metric that quantifies image quality degradation caused by processing such as data compression or by losses in data transmission. For image fusion, SSIM measures the structural similarity between source images and fused image from brightness, contrast and structure. Simplified SSIM is defined as follows:

$$SSIM(X, F) = \frac{(2\mu_x \mu_f + c_1)(2\sigma_{xf} + c_2)}{(\mu_x^2 + \mu_f^2 + c_1)(\sigma_x^2 + \sigma_f^2 + c_2)} \tag{5}$$

$$SSIM = \frac{SSIM(V, F) + SSIM(I, F)}{2} \tag{6}$$

where $SSIM(X, F)$ denotes the structural similarity between source image X and fused image Y, $\mu_x$ and $\mu_f$ denote the mean value of the image, $\sigma_x$ and $\sigma_f$ denote the standard deviation, $\sigma_{xf}$ is the standard covariance correlation of two images, $c_1$ and $c_2$ are constants to keep the denominator from being 0. $SSIM(V, F)$ and $SSIM(I, F)$ denote the structural similarities between visible/infrared image and fused image. The larger the SSIM value, the better the fusion effect.

$Q_{abf}$ [2] is a quality index which gives an indication of how much of the salient information contained in each of the input images has been transferred into the fused image without introducing distortions. $q_abf$ is defined as follow:

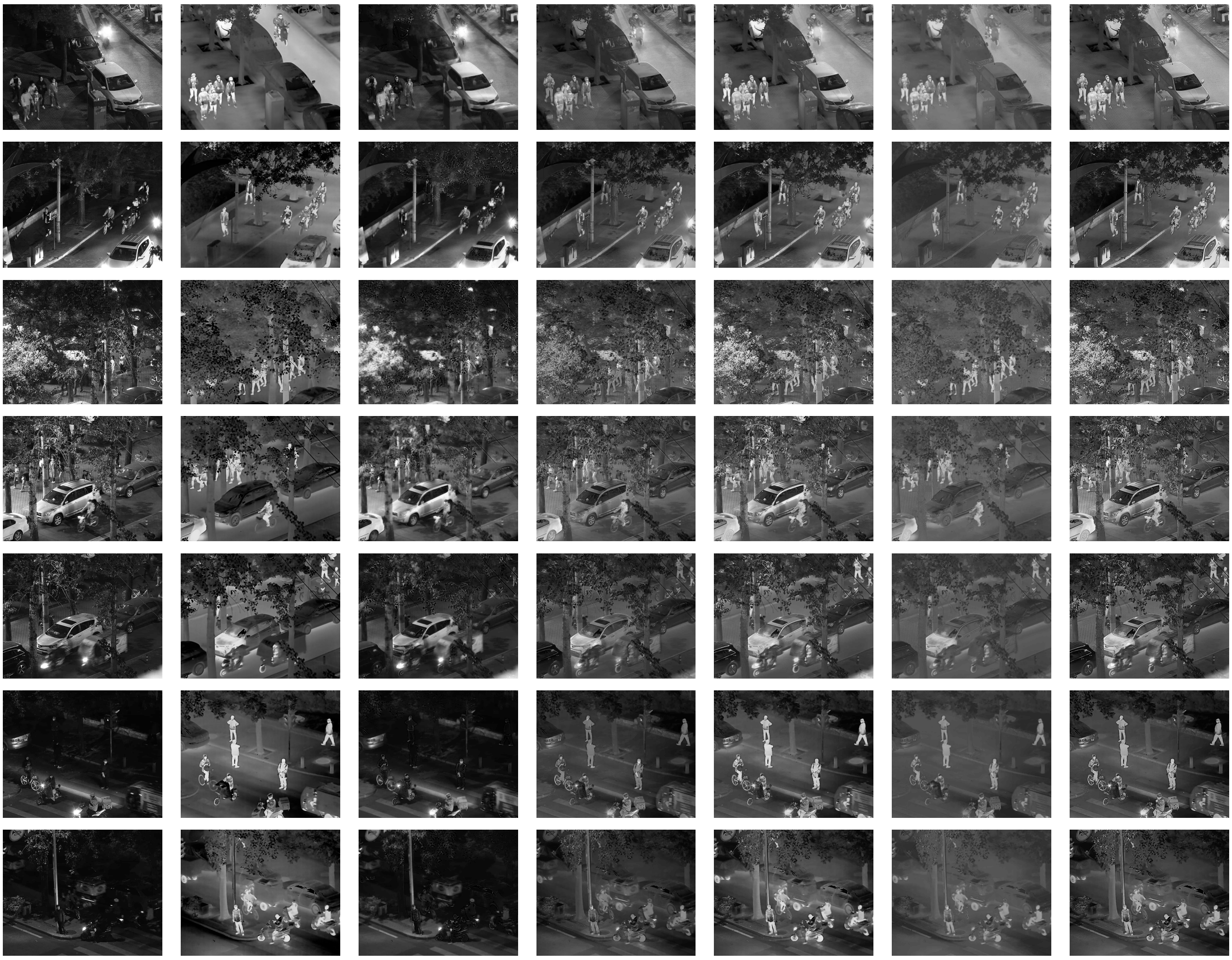
$$Q(a,b,f) = \frac{1}{|W|}\sum_{\omega\in W}(\lambda(\omega)Q_0(a,f|\omega) + (1-\lambda(\omega))Q_0(b,f|\omega)) \tag{7}$$

where W is the family of all windows and $|W|$ is the cardinality of W, $Q_0(a,f|\omega)$ is a measure for the similarity of the vectors x and y in the sliding window $\omega$ and takes values between -1 and 1. The higher the value of $Q_abf$, the better the quality of the fused image.

VIFF [1] utilizes the models in VIF to capture visual information from the two source fused pairs. With the help of an effective visual information index, VIFF measures the effective visual information of the fusion in images, while "effective visual information" is defined as the maximum visual information of all the source-fused image pairs.

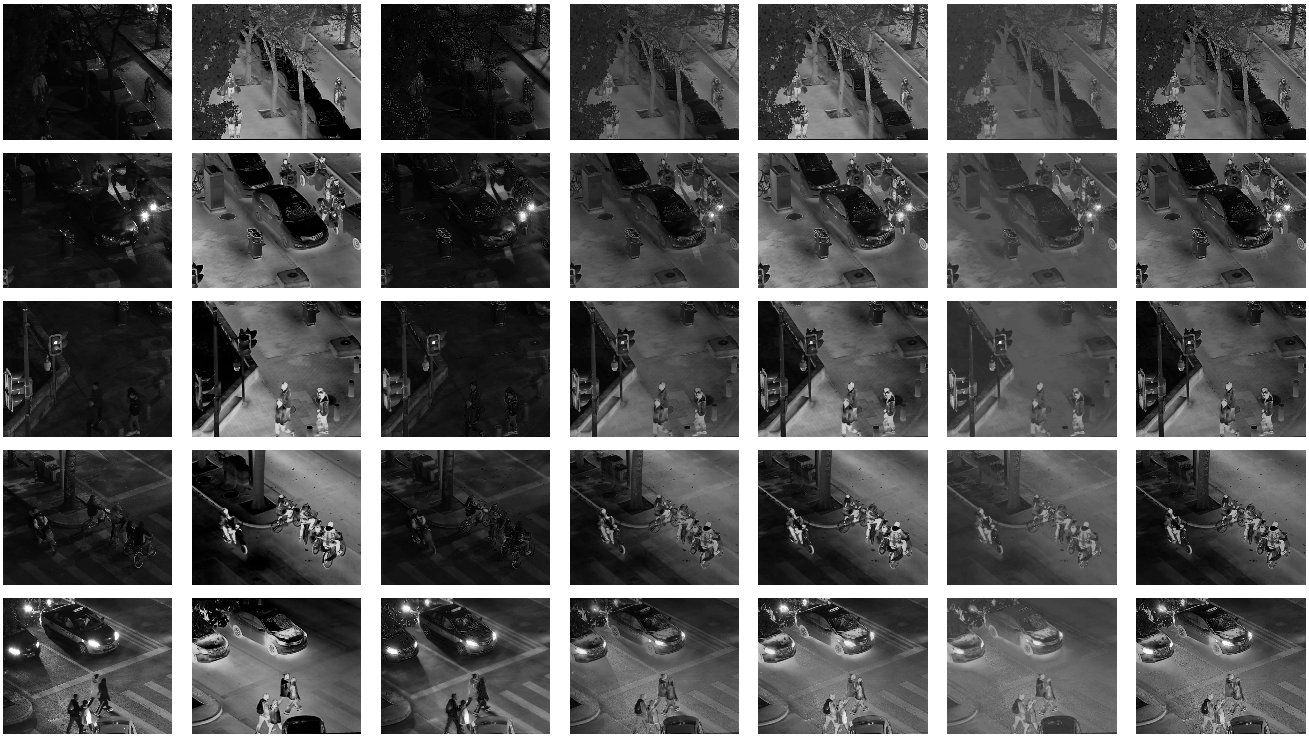
## 2. More Examples of LLVIP and Their Fusion Results

More examples of our LLVIP dataset and their fusion results are shown in Fig. 1 and Fig. 2.



(a) visible (b) infrared (c) GTF (d) densefuse_add (e) densefuse_$l_1$ (f) FusionGAN (g) IFCNN

Figure 1: More examples (part 1) of LLVIP and fusion results of several fusion algorithms on the LLVIP dataset. From left to right: (a) visible images, (b) infrared images, (c) GTF results, (d) densefuse_add results, (e) densefuse_$l_1$ results, (f) FusionGAN results, (g) IFCNN results.



(a) visible (b) infrared (c) GTF (d) densefuse_add (e) densefuse_$l_1$ (f) FusionGAN (g) IFCNN

Figure 2: More examples (part 2) of LLVIP and fusion results of several fusion algorithms on the LLVIP dataset. From left to right: (a) visible images, (b) infrared images, (c) GTF results, (d) densefuse_add results, (e) densefuse_$l_1$ results, (f) FusionGAN results, (g) IFCNN results.

# References


[1] Yu Han, Yunze Cai, Yin Cao, and Xiaoming Xu. A new image fusion performance metric based on visual information fidelity. *Information fusion*, 14(2):127–135, 2013. 1, 2

[2] Gemma Piella and Henk Heijmans. A new quality metric for image fusion. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 3, pages III–173. IEEE, 2003. 1, 2

[3] Guihong Qu, Dali Zhang, and Pingfan Yan. Information measure for performance of image fusion. *Electronics letters*, 38(7):313–315, 2002. 1

[4] Chaveli Ramesh and T Ranjith. Fusion performance measures and a lifting wavelet transform based algorithm for image fusion. In *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002.(IEEE Cat. No. 02EX5997)*, volume 1, pages 317–320. IEEE, 2002. 1

[5] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 1