Robust Perceptual Night Vision in Thermal Colorization

Feras Almasri[a] and Olivier Debeir[b]
LISA - Laboratory of Image Synthesis and Analysis, Université Libre de Bruxelles
CPI 165/57, Avenue Franklin Roosevelt 50, 1050 Brussels, Belgium

Keywords: Colorization, Deep learning, Thermal images, Nigh Vision.

Abstract: Transforming a thermal infrared image into a robust perceptual colour visual image is an ill-posed problem due to the differences in their spectral domains and in the objects' representations. Objects appear in one spectrum but not necessarily in the other, and the thermal signature of a single object may have different colours in its visual representation. This makes a direct mapping from thermal to visual images impossible and necessitates a solution that preserves texture captured in the thermal spectrum while predicting the possible colour for certain objects. In this work, a deep learning method to map the thermal signature from the thermal image's spectrum to a visual representation in their low-frequency space is proposed. A pan-sharpening method is then used to merge the predicted low-frequency representation with the high-frequency representation extracted from the thermal image. The proposed model generates colour values consistent with the visual ground truth when the object does not vary much in its appearance and generates averaged grey values in other cases. The proposed method shows robust perceptual night vision images in preserving the object's appearance and image context compared with the existing state-of-the-art.

1 INTRODUCTION

Humans have reasonable night vision with poor capabilities given improper environments. They have poor vision in low light conditions but with the advantage of rich colour vision in better lighting conditions. Human eyes have cone photoreceptor cells which are colour perception sensitive and rod photoreceptor cells which are receptive to brightness. The cones are unable to adapt well in low lighting conditions.

Colour vision is very important to the human brain. It helps to identify objects and to understand the surrounding environment. Studies (Cavanillas, 1999) (Sampson, 1996) have shown that human brain interpretation with colour vision improves the accuracy and the speed of object detection and recognition as compared to monochrome or false-colour visions. Due to this biologically limited interpretability, artificial night vision has become increasingly important in military missions, pharmaceutical studies, driving in darkness, and in security systems.

The use of thermal infrared cameras has seen an important increase in many applications, due to their long wavelength which allows capturing the objects invisible heat radiation despite lighting conditions. They are robust against some obstacles and illumination variations and can capture objects in total darkness. However, the human visual interpretability of thermal infrared images is limited, and so transforming thermal infrared images to visual spectrum images is extremely important.

The mapping process from monochrome visual images into colour images is called colorization, which has been broadly investigated in computer vision and image processing (Isola et al., 2017) (Zhang et al., 2016) (Larsson et al., 2016) (Guadarrama et al., 2017). However, it is an ill-posed problem because the two images are not directly correlated. A single object in the grayscale domain has a single representation while it might have different possible colour values in its true colour image counterpart. This is also true in the thermal images with additional challenging problems. For instance, a single object with different temperature conditions will have different thermal signature that can correspond to a single-colour value, while the thermal signature of two identical material objects at the same temperature conditions will look identical in the thermal infrared images, but have different colour values in their visual image counterpart.

[a] https://orcid.org/0000-0001-9321-6828
[b] https://orcid.org/0000-0002-6461-1551

Almasri, F. and Debeir, O.
Robust Perceptual Night Vision in Thermal Colorization.
DOI: 10.5220/0008979603480356
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 4: VISAPP*, pages 348-356
ISBN: 978-989-758-402-2; ISSN: 2184-4321

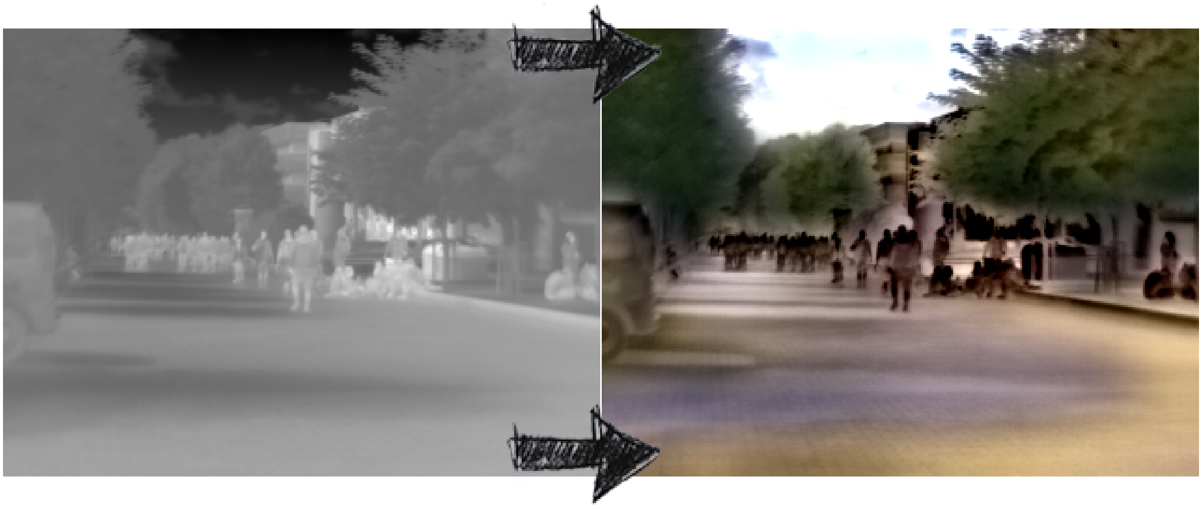


Figure 1: An example of mapping a thermal image to a color visual image is presented. (left): a thermal image from the ULB17-VT.V2 test set, and, (right): its colorized counterpart. This approach generates color values consistent with the color visual ground truth and preserves objects' textures from the thermal representation.

Transforming thermal infrared images to visual images is a very challenging task since they do not have the same electromagnetic spectrums and so their representations are different. In grayscale image colorization, the problem is to transform the luminance values into only the chrominance values, while in thermal image colorization, the problem requires estimating the luminance and the chrominance given only the thermal signature. Accordingly, a delivered solution should consider all of these challenges and also provide a method for preserving the representation of the objects in the thermal spectrum, while predicting the possible colour of known relatively fixed in space and time objects, such as the sky, tree leaves, street, traffic signs.

This paper addresses the problem of transforming the thermal images to consistent perceptual visual images using deep learning models. Our method predicts the low-frequency information of the visual spectrum images and preserves the high-frequency information from the thermal infrared images. A pansharpening method is then used to merge these two bands and creates a plausible visual image.

2 RELATED WORKS

Earlier grayscale image colorization required human guidance to manually apply colour strokes to a selected region or to give a reference image with the same colour palette. This should help the model to assume the similar neighborhood intensity values and assign them a similar color, e.g. Scribble (Levin et al., 2004), or Similar images (Welsh et al., 2002), (Ironi et al., 2005). Recently, the successful applications of convolutional neural networks (ConvNet) have encouraged researchers to investigate automatic end-to-end ConvNet based model on the grayscale colorization problem (Cao et al., 2017), (Iizuka et al., 2016), (Cheng et al., 2015), (Guadarrama et al., 2017).

A few researchers have investigated the colorization of near-infrared images (NIR) (Zhang et al., 2018), (Limmer and Lensch, 2016) and have shown a high performance, due to the high correlation between the NIR and RGB images. Their two wavelengths differ only slightly in the red spectrum and thus they have similar visual light representation correlated in the red channel. In contrast, thermal images taken from the long-wavelength infrared spectrum (LWI) do not correlate with the visual images since they are measured by the emitted radiation linked to the objects' temperature. Therefore, predicting the colour of an object in its thermal signature requires a local and global understanding of the image context.

Recently Berg et al. (Berg et al., 2018) and Nyberg et al. (Nyberg, 2018) presented a fully automatic ConvNet on a thermal infrared to RGB image colorization problem using different objective functions. Their models illustrated a robust method against image pair misalignment. However, the generated images suffer from a high blur effect and artefacts in different locations in the images, e.g. missing objects from the scene, object deformations and some failure images. Kuang et al. in (Kuang et al., 2018) used a conditional generative adversarial loss to generate a realistic visual image, with the perceptual loss based on the VGG-16 model, the TV loss to ensure spatial smoothness, and the MSE as content loss. Their work presented better realistic colour representations with fine details but also suffered from the same artefacts, missing objects and object deformations.

The previous works were trained on the KAIST-MS dataset (Hwang et al., 2015) which consists of 95,000 thermal-visual images captured from a device mounted on a moving vehicle. Images were captured during day and night by a thermal camera with an output size of 320x256 and interpolated to have the same size as the visual images (640x512) using an unknown method and normalized using an unknown histogram equalization method. The procedure used to train the models in previous work reduces the size of the thermal images to their original size and then trains the models only on day time images. The frames were extracted from the video sequence, so it should be considered that, several subsequent images are very similar in most of the sets and it is possible to overfit the dataset. It is also possible that the equalization coupled with the rescaling methods changed the thermal value distribution. Therefore, the proposed model is also trained on the ULB17-VT dataset (Almasri and Debeir, 2018) which contains raw thermal images.

3 METHOD

For this work, the target is to transform the thermal infrared images from their temperature representations to colour images. For this reason, this work builds on existing works that have looked at the thermal colorization problem and uses the proposed network architecture by Berg et al. (Berg et al., 2018) with small modifications adapted to our outputs.

Preprocessing steps are assumed necessary when the ULB17-VT dataset is used. Images are normalized to $[0-1]$ using instance normalization in contrast with the KAIST-MS dataset which used histogram equalization. Spikes that occur with sharp low/high temperatures are detected and smoothed using a convolution kernel.

The method proposed here is to transform the thermal image to low-frequency (LF) information in the colour visual image space in a match with the LF information in the ground truth visual image. The final colourized image is acquired by applying a post-processing pansharpening step. This process is done by merging the predicted visual LF information with the high-frequency (HF) information extracted from the input thermal image. This step is assumed necessary to maintain an object's appearance from the thermal signature and to preserve it in the predicted colourized images. It also helps avoid high artefact occurrences when object representations are different between the two spectrums.

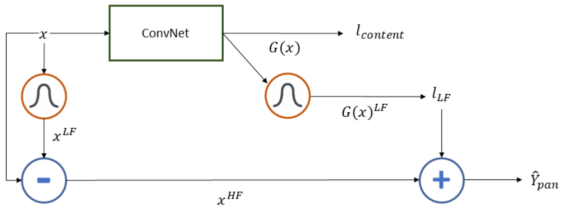


Figure 2: Proposed Model. Model (G) in orange is the Gaussian layer.

3.1 Proposed Model

The proposed model, as illustrated in Fig. 2, takes the thermal image as input and generates a fully colourized visual image. For this generated output, L1 content loss $l_{content}$ is used as an objective function to measure the dissimilarities with the ground truth visual image. The low-frequency information is then obtained from the generated colourized image $G(x)^{LF}$ and from the ground truth visual image Y^{LF} by applying a Gaussian convolution layer with a kernel of width 25 x 25 and $\sigma = 12$. The dissimilarities between the LF information of the two images is measured using the objective function l_{lf} which is the MSE loss. The total loss is a weighted sum of the L1 and MSE multiplied by $\alpha = 10$ since the MSE loss value is smaller than L1.

$$l_{total} = l_{content} + \alpha \cdot l_{lf} \tag{1}$$

3.2 Representation and Pre-processing

The pansharpening method is used as shown in Fig. 2 as a final post-processing step. The thermal low-frequency information x^{LF} is first obtained by applying a Gaussian layer on x. The thermal high-frequency information x^{HF} is then extracted by subtracting x^{LF} from x. The thermal image is represented with three channels in order to add them to the visual RGB images. The final colourized thermal image $\hat{Y}_{pan}$ is obtained by adding the input x^{HF} weighted by λ to the generated low-frequency information $G(x)^{LF}$ as:

$$\hat{Y}_{pan} = G(x)^{LF} + \lambda x^{HF} \tag{2}$$

The pansharpening method is first applied on the ground truth visual images to experience and visualize the pan-sharped colourized images before training the model. The thermal signature of the sky in the thermal images is very low with respect to other objects, while humans and other heated objects have a higher thermal signature. The normalization process

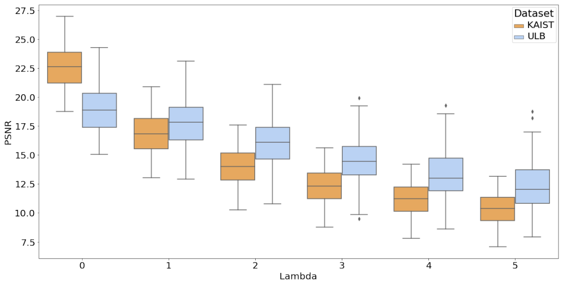
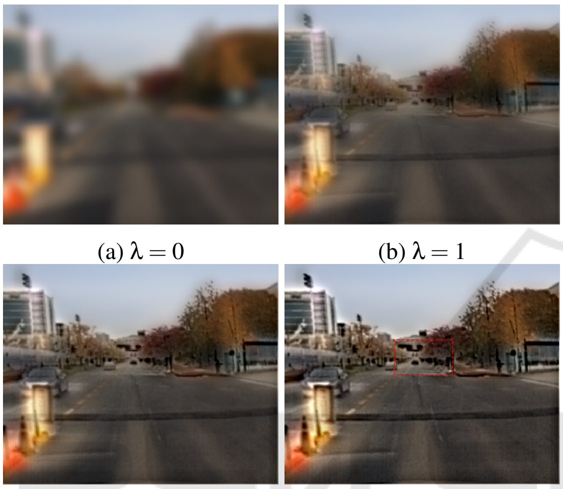


Figure 3: Boxplot of PSNR for $\lambda = 0,1,2,3,4,5$ on ULB17-VT.V2 test set and on KAIST-MS set00-V000 set.



(a) $\lambda = 0$ (b) $\lambda = 1$

(c) $\lambda = 2$ (d) $\lambda = 3$

Figure 4: Pansharpening visualization from KAIST-MS dataset on S6V0I00000 with $\lambda = 0,1,2,3$.

makes the sky values very close to zero, while in the visual images this value should be around one. For this reason, the thermal infrared images are inverted before any processing which results in a value around one for the sky in the thermal images.

The proposed method relies on maintaining the high-frequency information taken from the thermal images, as this can reduce the evaluation results compared to the state-of-the-art when the pixel-wise measurement is used. For validation purposes, the PSNR between $\hat{Y}_{pan}$ and y with $\lambda = 0,1,2,3,4,5$ was measured as shown in Fig. 3. This gives an idea of the maximum validation value that can be achieved using the proposed model. The synthesised images are represented as a perceptual visualization quality as shown in Fig. 4. The value $\lambda = 3$ was chosen as a trade-off between better perceptual image quality and a reasonable PSNR with the average of 14.5 for ULB17-VT.V2 and 12.31 for KAIST-MS. If λ is decreased the PSNR value increases, but with less plausible perceptual images.

When the weighted thermal HF information is added to the visual LF information, the synthesized image could have values out of the band $[0-1]$ in some areas. This results in a black or white color effect when the image is clipped to the range $[0-1]$ as shown in the red rectangle in Fig. 4. Re-normalizing the image instead of clipping can reduce the image contrast or affect the true colour values since the low frequency information on the three RGB channels is being obtained and added. This problem can be solved by exploring different normalization methods in the pre-processing step and different merging procedures in the post-processing step.

De-spiked thermal images are obtained using a convolution kernel of width 5 x 5, which replaces the centre pixel with the median value if the pixel value is three times greater than the standard deviation of the kernel area.

3.3 Networks Architecture

The network architecture proposed in (Berg et al., 2018) from their repository was used. [1]. Two models were trained as follows:

- TICPan-Bn The proposed method using the network architecture in (Berg et al., 2018).
- TICPan The proposed method using the same network architecture, and replacing the batch normalization layer with the instance normalization layer. It shows better enhancement in colour representations and in the metric evaluations.

4 EXPERIMENTS

4.1 Dataset

For this work the ULB17-VT dataset (Almasri and Debeir, 2018) which contains 404 visual-thermal image pairs was used. The number of images was increased to 749 visual-thermal images using the same device and 74 pairs were held for testing. Thermal images were extracted in their raw format and logged in with 16-bit float per-pixel. This new dataset, ULB-VT.v2, is available on [2].

The KAIST-MS dataset (Hwang et al., 2015) was also used and the exiting works on thermal colorization problem were followed. Training was only done on day time images and resized the thermal images to their original resolution of 320 x 256 pixels. The images in KAIST-MS were recorded continuously during driving and stopping the car. This results in a high

[1] https://github.com/amandaberg/TIRcolorization

[2] http://doi.org/10.5281/zenodo.3578267

number of redundant images and explains the overfitting behaviour and the failure results in previous work. For this reason, only every third image is taken in the training set to yield a set with 10,027 image pairs, while all of the images in the test set are used.

4.2 Training Setup

All experiments were implemented in Pytorch and performed on an NVIDIA TITAN XP graphics card. TIR2Lab (Berg et al., 2018) and TIC-CGAN (Kuang et al., 2018) were re-implemented and trained as explained in the original papers.

The proposed model, TICPan, trained using ADAM optimizer with default Pytorch parameters and weights were initialized with He normal initialization (He et al., 2015). All experiments were trained for 1000 epochs and the learning rate was initialized with $8e^{-4}$ with decay after 400 epochs. The LeakyReLU layers parameter was set to $\alpha = 0.2$ and the dropout layer was set to 0.5.

In each training batch, 32 cropped images of size 160 x 160 were randomly extracted. For each iteration, a random augmentation was applied by flipping horizontally or vertically and rotating in the $[-90°, 90°]$. Since the number of training images in KAIST-Ms is 14 times more than ULV-VT.v2, the number of iterations for the model to train on the ULV-VT.v2 was increased to match the model trained on KAIST-MS.

For validation, the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and root-mean-square error (RMSE) were used between the generated colorized images and the true images.

4.3 Quantitative Evaluation

The proposed model was evaluated on transforming thermal infrared images to RGB images compared with the state-of-the-art using the measurement metrics shown in Table 1.

The proposed model evaluation was performed on the full colorized thermal image, which is the result of the fusion of the predicted visual LF information and the input thermal HF information. This resulted in a higher pixel-wise error compared to other models since the HF content of the image was taken from the thermal domain. However, our method achieved comparable results with the synthesized images as shown in Fig. 3.

It is believed that the pixel-wise metrics are not suitable for the colorization problem where the perception of the image has an important role. The TIR2Lab achieved higher evaluation values while their generated images are uninterruptable. TIC-CGAN has 12.266 million parameters that explain the overfitting behaviour in its generated images. TICPan-BN was excluded because it has the lowest evaluation values and less comparable quality images.

4.4 Qualitative Evaluation

Four examples are presented in Fig. 8 on the ULB17-VT.v2 dataset. The TIR2Lab model generated approximated good colour representations for trees with blur effect but failed to produce fine textures and to preserve the image content. On the hand, the TIC-CGAN model generated better image colour quality with fine textures and were more realistic. This is very recognizable, as an over-fitting behaviour, when the test image comes from the same distribution as the densely represented images in the training set such as image number (650).

TICPan generates images that have strong true colour values for objects that are relatively fixed in space and time, such as sky, tree leaves, and streets and buildings. Sky is represented in white or light blue colour, trees are in different shades of green, and streets and buildings also represented with approximated true colour values. However, objects like humans are represented in grey or in black due to the clipping effect. Our method assures that the object thermal signature does not disappear in image transformation or get deformed. The model cannot predict true colour values for the varying objects but it predicts an averaged colour value represented in grey and the final pansharpening process maintains their appearance in the generated colourized images.

In Fig. 9 four examples are presented on the KAIST-MS dataset. The TIR2Lab method produced approximate good true chrominance values but it has heavily blurred images and suffers from recovering fine textures accurately. The produced artefacts are very obvious in the generated images and some objects, such as the walking person in (S6V3I03016) are missing in their outputs. The TIC-CGAN model produced better perceptual colourized thermal images with realistic textures and fine details, but they suffer from the same countereffects of missing objects and objects deformation. This is due to the use of GAN adversarial loss which learns the dataset distribution and estimates what should appear in each location, and also because of the large size of the model and its over-fitting behaviour. This is seen in (S8V2I01723) in the falsely generated road surface markings and in the missing person in (S6V3I03016). In contrast, the proposed TICPan model does not generate very plausible colour values in the KAIST-MS dataset but it

Table 1: Average evaluation results on 74 images in ULB-VT version 2 dataset and 29,179 images in KAIST-MS dataset.

Model	Parameters	Dataset	PSNR	SSIM	RMSE
TIR2Lab	1.46M	ULB-VT.V2 KAIST-MS	14.404 14.090	0.335 0.565	0.194 0.204
TIC-CGAN	12.266M	ULB-VT.V2 KAIST-MS	15.475 16.010	0.313 0.552	0.174 0.165
TIC-Pan-BN	1.46M	ULB-VT.V2 KAIST-MS	12.559 12.944	0.215 0.373	0.239 0.228
TIC-Pan	1.46M	ULB-VT.V2 KAIST-MS	13.078 13.922	0.228 0.404	0.226 0.205

generates robust perceptual night vision images that maintain objects' appearances.

4.4.1 Deformation and Missing Objects

Fig. 9 shows missing objects in the TIC-CGAN generated images, such as the person in (S0V0I00601) and the cars in (S0V0I01335). We can also recognize the object deformation in image number (428) and image number (598), while in the TICPan model objects are retained in the generated images.

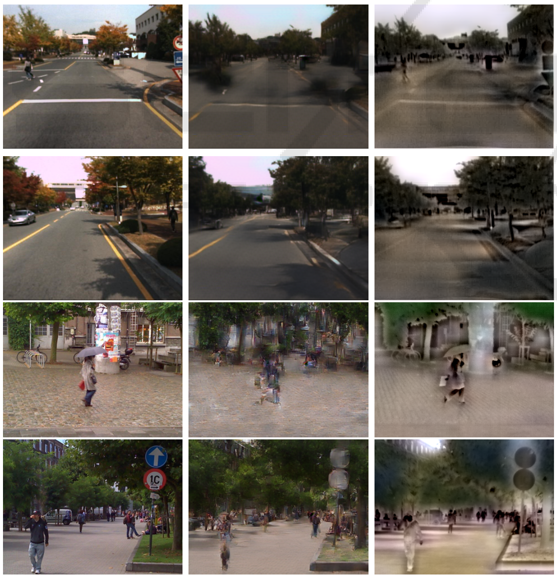


Figure 5: From left to right: True RGB, TIC-CGAN and TICPan. From top to bottom: (S0V0I00601) and (S0V0I01335) form KAIST-MS and (428) and (598) from ULB-VT.v2.

4.4.2 Overfitting Behavior

Fig. 6 illustrates the over-fitting problem in the TIC-CGAN model. Because of its size, it has 12M parameters and is 12 times bigger than the proposed model. This makes it very easy for the model to overfit the dataset and not perform generalisation in the unseen data. In image number (1250), the model can predict the exact colour of the two cars because a similar image appeared in the training set. In the second image number (S0V0I00613), whenever an object comes from the left with a size similar to a bus, the model will predict it as a bus with red colour. The TICPan model cannot predict the exact colour of cars, but instead generates an average grey colour.

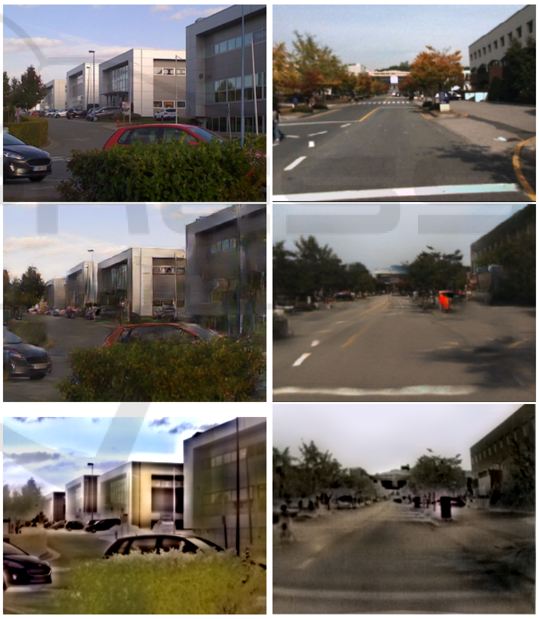


Figure 6: From left to right: (1250) from ULB-VT.v2 and (S0V0I00613) from KAIST-MS test set. From top to bottom: True RGB, TIC-CGAN and TICPan.

4.4.3 Night Vision

The TIC-CGAN model failed to generate interpretable images using images that were taken at night, because the image distribution and the image contrast were different from the training images. However, the TICPan model does not suffer from this failure thanks to the pansharpenning process as shown in Fig. 7. In image number (1784), the true RGB image is completely dark and the TICPan model generates a robust perceptual night vision image as compared to the TIC-CGAN model. This is also illustrated in image number (S9V0I00000), where the TICPan model generates a night vision image with less artefacts than the TIC-CGAN model. It should be noted that these artefacts are due to the histogram equalization method used in KAIST-MS.

5 CONCLUSIONS

The objective in this study was to address the problem of transforming thermal infrared images to visual images with robust perceptual night vision quality. In contrast to the existing methods that map images automatically from their thermal signature to chrominance information, our proposed model seeks to maintain the appearance of objects in their thermal representation from the thermal images and to predict possible colour values.

The evaluation showed that the proposed model has better perceptual images with fewer artefacts and the best representation for night images. This confirms the model generalization capability. The generated images are robust and reliable enabling users to better interpret the images while using night vision. For objects or cases in which missing or deformed objects can cause dramatic accidents, the pan sharpening process is of critical necessity.

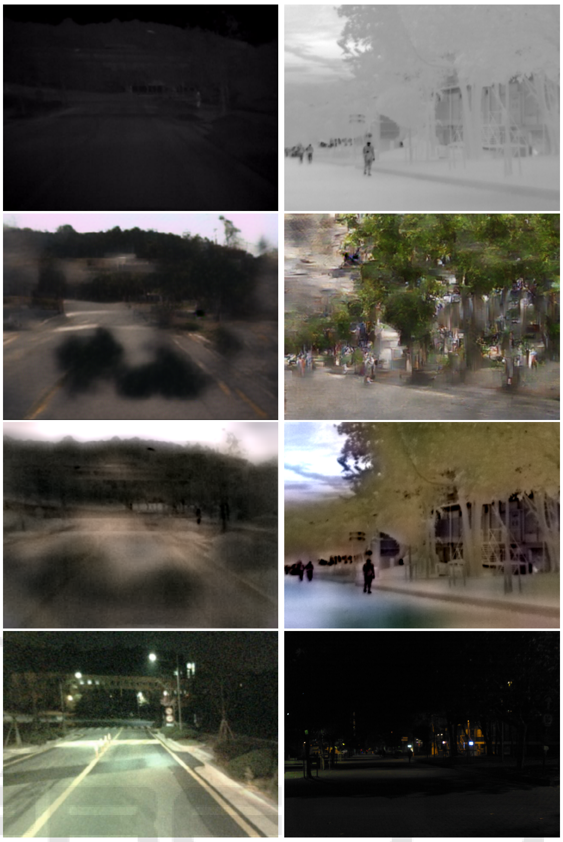


Figure 7: From top to bottom: Thermal image, TIC-CGAN and TICPan. Left (S9V0I00000) from KAIST-MS and right (1784) ULB-VT.v2.

ACKNOWLEDGEMENTS

This work was supported by the European Regional Development Fund (ERDF) and the Brussels-Capital Region within the framework of the Operational Programme 2014-2020 through the ERDF-2020 project F11-08 ICITY-RDI.BRU. We thank Thermal Focus BVBA for their support.

REFERENCES

Almasri, F. and Debeir, O. (2018). Multimodal sensor fusion in single thermal image super-resolution. *arXiv preprint arXiv:1812.09276*.

Berg, A., Ahlberg, J., and Felsberg, M. (2018). Generating visible spectrum images from thermal infrared. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1143–1152.

Cao, Y., Zhou, Z., Zhang, W., and Yu, Y. (2017). Unsupervised diverse colorization via generative adversarial networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 151–166. Springer.

Cavanillas, J. A. A. (1999). The role of color and false color in object recognition with degraded and non-degraded images. Technical report, NAVAL POSTGRADUATE SCHOOL MONTEREY CA.

Cheng, Z., Yang, Q., and Sheng, B. (2015). Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423.

Guadarrama, S., Dahl, R., Bieber, D., Norouzi, M., Shlens, J., and Murphy, K. (2017). Pixcolor: Pixel recursive colorization. *arXiv preprint arXiv:1705.07208*.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034.

Hwang, S., Park, J., Kim, N., Choi, Y., and So Kweon, I. (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1037–1045.

Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2016). Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization

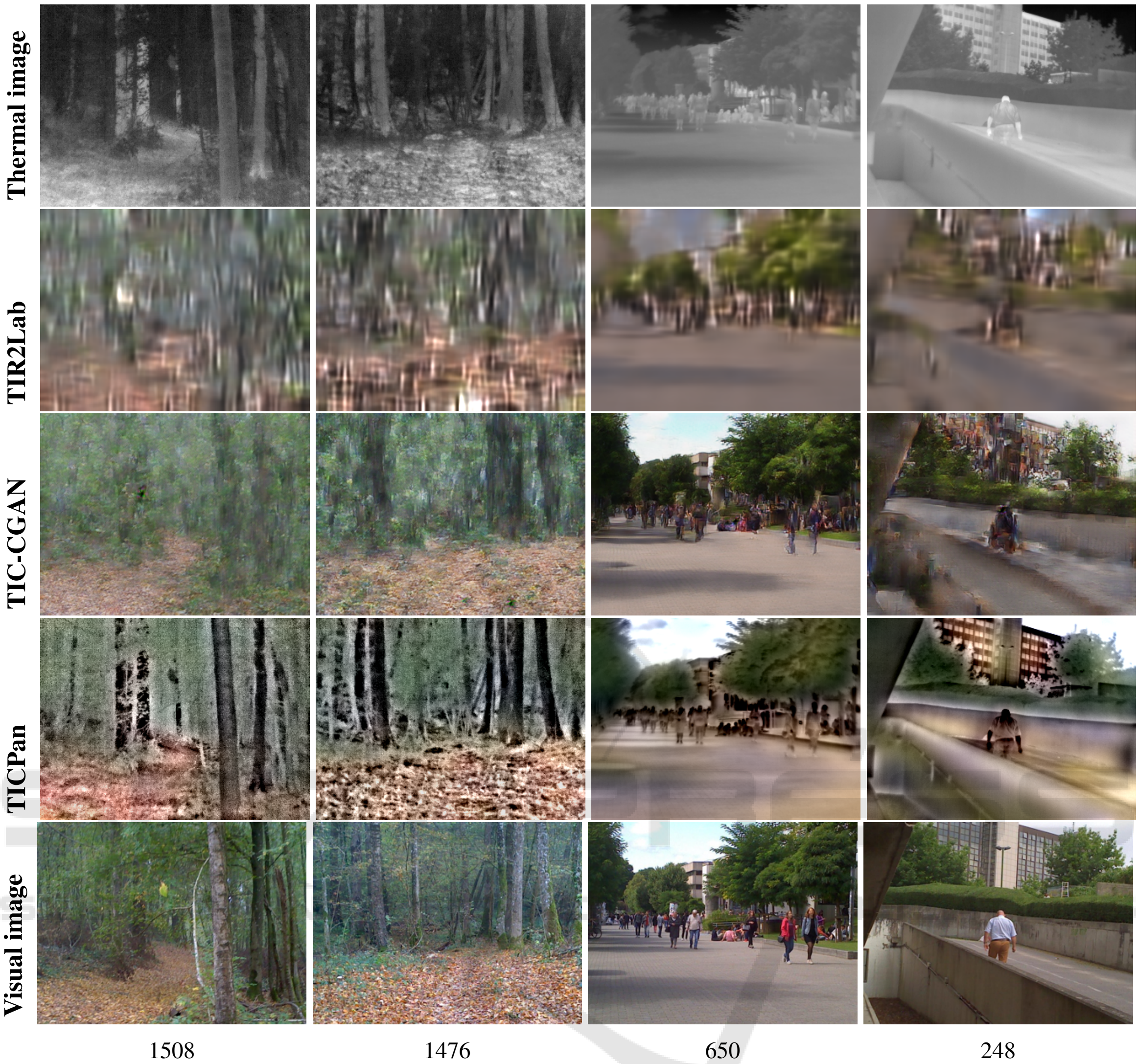


Figure 8: Examples of colorized results on ULB-VT.v2 test set. The numbers represent the image names.

with simultaneous classification. *ACM Transactions on Graphics (TOG)*, 35(4):110.

Ironi, R., Cohen-Or, D., and Lischinski, D. (2005). Colorization by example. In *Rendering Techniques*, pages 201–210. Citeseer.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Kuang, X., Sui, X., Liu, C., Liu, Y., Chen, Q., and Gu, G. (2018). Thermal infrared colorization via conditional generative adversarial network. *arXiv preprint arXiv:1810.05399*.

Larsson, G., Maire, M., and Shakhnarovich, G. (2016). Learning representations for automatic colorization. In *European Conference on Computer Vision*, pages 577–593. Springer.

Levin, A., Lischinski, D., and Weiss, Y. (2004). Colorization using optimization. In *ACM transactions on graphics (tog)*, volume 23, pages 689–694. ACM.

Limmer, M. and Lensch, H. P. (2016). Infrared colorization using deep convolutional neural networks. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 61–68. IEEE.

Nyberg, A. (2018). Transforming thermal images to visible spectrum images using deep learning.

Sampson, M. T. (1996). *An assessment of the impact of fused monochrome and fused color night vision displays on reaction time and accuracy in target detection*. PhD thesis, Monterey, California. Naval Postgraduate School.

Welsh, T., Ashikhmin, M., and Mueller, K. (2002). Transferring color to greyscale images. In *ACM transactions on graphics (TOG)*, volume 21, pages 277–280. ACM.

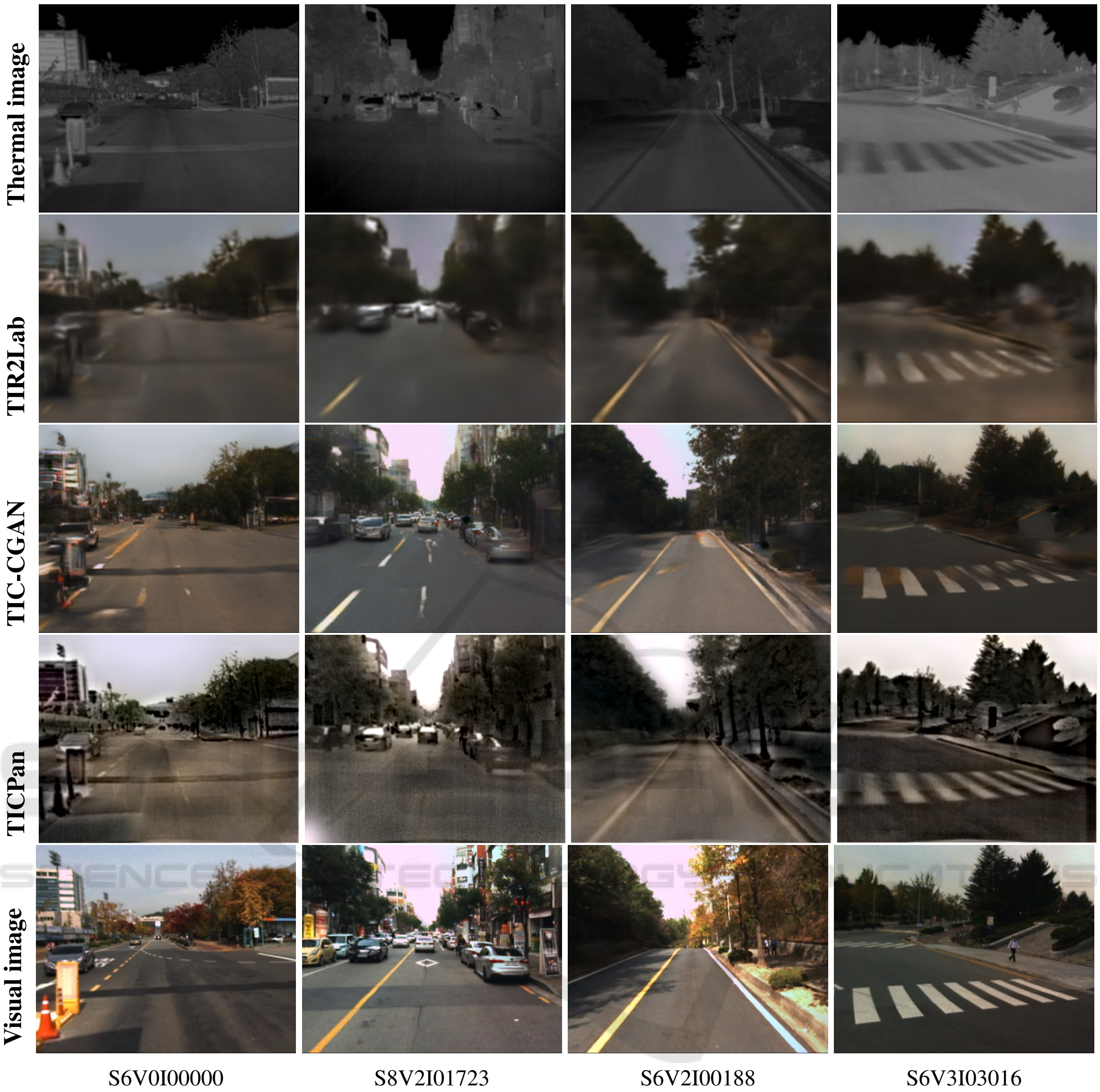


Figure 9: Examples of colorized results on KAIST-MS test set. The numbers represent the image place and their names.

Zhang, R., Isola, P., and Efros, A. A. (2016). Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer.

Zhang, T., Wiliem, A., Yang, S., and Lovell, B. (2018). Tv-gan: Generative adversarial network based thermal to visible face recognition. In *2018 International Conference on Biometrics (ICB)*, pages 174–181. IEEE.