Local estimation of parametric point spread functions in thermal images via convolutional neural networks

Florian Piras[a], Edouard De Moura Presa[b], Peter Wellig[b], and Michael Liebling[a,c]

[a]Idiap Research Institute, Martigny, Switzerland
[b]armasuisse Science and Technology, Thun, Switzerland
[c]University of California, Santa Barbara, CA, USA

ABSTRACT

Thermal image formation can be modeled as the convolution of an ideal image with a point spread function (PSF) that characterizes the optical degradations. Although simple space-invariant models are sufficient to model diffraction-limited optical systems, they cannot capture local variations that arise because of nonuniform blur. Such degradations are common when the depth of field is limited or when the scene involves motion. Although space-variant deconvolution methods exist, they often require knowledge of the local PSF. In this paper, we adapt a local PSF estimation method (based on a learning approach and initially designed for visible light microscopy) to thermal images. The architecture of our model uses a ResNet-34 convolutional neural network (CNN) that we trained on a large thermal image data set (CVC-14) that we split in training, tuning, and evaluation subsets. We annotated the sets by synthetically blurring sharp patches in the images with PSFs whose parameters covered a range of values, thereby producing pairs of sharp and blurred images, which could be used for supervised training and ground truth evaluation. We observe that our method is efficient at recovering PSFs when their width is larger than the size of a pixel. The estimation accuracy depends on the careful selection of training images that contain a wide range of spatial frequencies. In conclusion, while local PSF parameter estimation via a trained CNN can be efficient and versatile, it requires selecting a large and varied training data set. Local deconvolution methods for thermal images could benefit from our proposed PSF estimation method.

Keywords: thermal imaging, point spread function estimation, convolution neural networks, space-variant deconvolution

1. INTRODUCTION

Thermal image formation can be mathematically modelled as a convolution operation between an ideal image (with contrast given by temperature) and a point spread function (PSF), which captures the diffraction-limited optics of the imaging system and corresponds to the image produced by an infinitely small light source. Knowledge of the PSF associated to an image is useful, in particular, for applications that aim at restoring image detail, for example via image deconvolution algorithm.[1] Although the PSF could be modelled or measured through experimental calibrated of the camera, such processes can be time-consuming or impossible to carry out if only images are available.

The ability to obtain an estimate of the PSF given *only* the image itself is therefore particularly desirable.

Furthermore since image formation may not be uniform over the entire image, for example, when the subject appear blurry due to motion or because part of the image is not within the depth of field, different PSFs (for each region of the image) are required to accurately model image formation. In the latter situations, direct measurement or calibration would not be viable to determine local PSFs. Instead, image-based estimation offers a practical alternative.

Shajkofci and Liebling[2] recently described and approach to estimate the PSF in the case of visible light for optical microscopy based on trained convolutional neural network. Previous studies[3–6] had used similar methods to estimate the direction and magnitude of motion blur in visible images.

Further author information: florian.piras@idiap.ch

In this paper, we adapted the method in Ref.[2] such that is applicable to thermal images, specifically, we investigated the ability of networks trained on visible and various types of thermal images (containing both textured and flat areas or mostly textured areas) to generalize.

This paper is organized as follows. In section 2, we present the methods we used to locally estimate PSF parameters on visible and thermal images. In Section 3, we present and compare the results of different experiments performed on visible and thermal images. Finally, we discuss the results in Section 4 and conclude in Section 5.

2. METHODS

The objective of our method is to estimate the PSF on thermal images using a trained convolutional neural network. We construct the training dataset from a large database of thermal images, which we assume being ideal (no degradation). This would typically be the case for images taken of scenes consisting of natural or human-made objects imaged in ideal conditions (steady camera, objects within the camera's depth of field, and well adjusted focus). We next pair each ideal image with a synthetic PSF (blur parameter, characterized e.g. by its full width at half maximum in units of pixels on the image) and combine them, via convolution, to form a synthetically blurred training image (Figure 1). An annotated training image therefore consists of the (synthetically) blurred image and the PSF parameter(s) used to produce the blur.

We split each image data set that we considered (described below) into two sets, one for trainingand another for evaluation (testing), which contained 90% and 10% of the images, respectively.

More specifically, we took large images (640 by 471 pixels) that we divided into patches of size 128 by 128 pixels, which we convolved with a PSF whose width parameter was randomly chosen within the range 0 - 9 (zero corresponds to not applying any blur degradation). The parameter corresponds to the full width at half maximum (FWHM) of a Gaussian. We added Poisson noise on the training set to improve the robustness of the next step,[2] training of a neural network: the degraded patches are fed into a Resnet34 network (modified for grayscale images input[2]) which locally estimates PSF parameters in the patch. We used moving windows that overlap by half the window (64 pixels). To use the model, a full size blurred image is selected and then divided into sub-images of predefined size (same size as the window used to generate the training set) with an overlap of half the window. These sub-images are grouped into a vector which constitutes the input of the model. The output of the model is the parameter estimate for all tiles without taking into account the overlap. To determine the parameter in the overlapping areas, we perform an interpolation by applying a median filter and local mean.

3. RESULTS

3.1 Training a PSF estimation model on visible images (Model(V))

In order to establish a point of comparison for the specific implementation and parameters of our method, we first trained the network with visible images. For this, we used the Places205[7] database. This database consists of 2.5 million images of scenes from 205 categories (such as abbeys, planes, fields, etc.). A selection of images from this database is shown in Fig. 2 and shows the diversity of the images. The image size is 256 by 256 pixels. We extracted 128 by 128 pixels windows from these images. The windows were overlapping by 64 pixels so that we could extract 9 sub-images per full image. 5000 images are selected randomly in the database. Each sub-image is blurred with a Gaussian blur that is characterized by a FWHM which is chosen at random (uniformly) in the range 0 to 9 pixels. 90% of the images are selected for the training set and the other 10% for the evaluation set. The network used is a modified ResNet34 that takes a grayscale image as input and outputs an estimate of the local blur parameter.[2] Figure 5a shows the evolution of the loss function. After about 30 epochs, good estimation is achieved. At Epoch 34, the loss functions are equal to 0.037 for training and 0.289 for evaluation. We denote the PSF estimation model trained on visible images as *Model (V)*. To estimate the performance of the trained model, we blurred several images with a Gaussian of known FWMH and compared the latter value with that estimated by the system. For this purpose, we selected 10 new images from the original database and FWMH values in the range 0 to 9. Figure 6a shows the average of the estimation parameter over the 10 blurred images (see curve that corresponds to the model trained on visible images, *Model (V)*) gives a good estimate of the parameters, except when the target parameter is equal to 0 where the estimated parameter is close to 1.

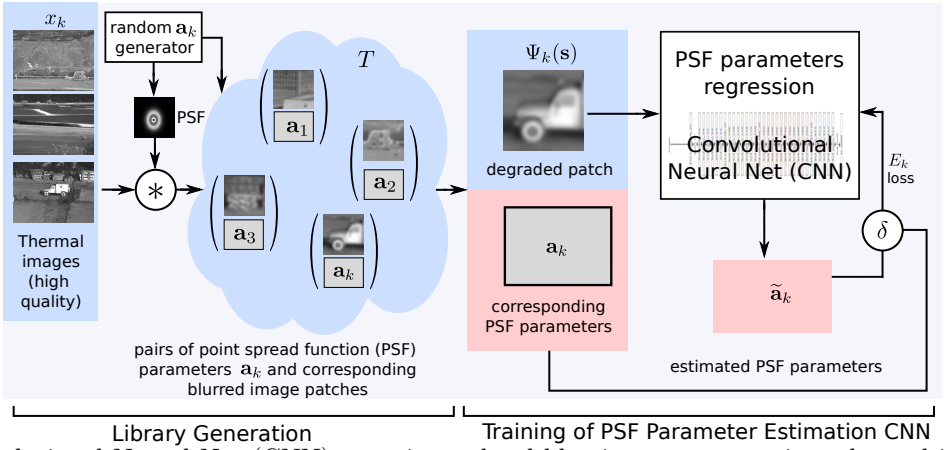


Figure 1: Convolutional Neural Net (CNN) to estimate local blurring parameters in a thermal image (adapted for thermal imaging from Ref.[2]). This approach has two steps. First, we generate a library of blurred patches from high-quality thermal images. Each patch is blurred with a randomly generated set of parameters to form an annotated image (blurred image, blurring parameters). The degraded patch is presented as an input to a CNN that estimates the parameters; during the learning phase, the CNN weights are adjusted by comparing the estimated parameters with the ground truth values from the training set.

3.2 Training a PSF estimation model on thermal images (Model(Th$_{t+f}$))

We next focused on training a model suitable for estimating the local PSF in thermal images. There are only few thermal image datasets that are widely available for research purposes. We used the CVC-14[8] dataset, which is composed of a few thousand thermal images of urban scenes taken during the day and night from a vehicle (examples shown in Fig. 3). The size of the images is 640 by 471 pixels. Similarly to the visible case, we set a window size of 128 by 128 pixels to extract example sub-images. Given the larger size of the thermal images (compared to the visible images from Places205) we could obtain a comparable number of example sub-images despite having fewer images in the thermal database. We selected 1000 images at random from the complete database (day and night images with and without people). We then blurred each sub-image with a Gaussian window, again with randomly chosen FWHM parameters in the range 0 to 9. As with the previous dataset, we also divided the images into a training set containing 90% of the selected images and an evaluation set that contained the remaining 10% of the images. We added Poisson noise to the training set sub-images. We carried out the training over 60 epochs. Figure 5b shows that the learning process appears to converge (the training loss at epoch 60 is 0.027) but the test loss remains high over the epochs (around 6). We denote the PSF estimation model on the thermal images as Model(Th$_{t+f}$), where $t+f$ refers to the fact that the images contain both flat and textured areas.

To investigate the performance of the trained model on larger images, we blurred 10 new images with a Gaussians of known FWMH (same FWHM in all areas of the image), then locally estimated the FWHM in all regions of the image, averaged the estimated values obtained in all regions and all images and finally compared that result with the actual (known) blurring parameter.

The curve Model(Th$_{t+f}$) in Figure 6b, shows the estimated parameter as a function of the expected parameter. We notice that the parameter estimate does not match the ground truth, in particular when the PSF parameter is low. It also performs only sligthly better than if the model trained on visible images (Model(V)) is used to estimate the PSF in thermal images.

3.3 Local performance variations

In order to investigate the lower performance we observed when training and evaluating a model on thermal images (Model(Th$_{t+f}$)), we visually inspected the performance on various example images. The left (Model (V)) and central (Model(Th$_{t+f}$)) columns in Fig. 7 show the local performance on three example images and


Figure 2: Preview visible dataset places 205

three PSF blurring widths (FWHM equal to 2, 6, or 9 pixels). While the local PSF estimation maps for the model trained on visible images (Model(V), left column), confirms the good global estimation performance of that model (for visible images), the local PSF estimation map for the model trained on thermal images (Model(Th_{t+f}), central column), shows large local variations in the estimation performance. In particular, we noticed that the performance is particularly poor in the lower third of most images, which exhibit little texture and are predominantly flat (Fig. 4).

3.4 Training a PSF estimation model on thermal images with flat areas removed (Model(Th_t))

We hypothesized that the poor performance of the model trained on thermal images (Model(Th_{t+f})), might be due to the fact that the lack of texture in the lower third of all thermal images in the database might bias the model toward over-estimating the PSF width.

To mitigate the influence that the lack of texture in the full thermal images (Th_{t+f}) might have, we trained a new model based on only the two upper thirds of the Th_{t+f} images in the CVC-14 database. We denote that model as Model(Th_t), where the t (without f) refers to the fact that the images are much more textured. Since these images cover only 2/3 of the original images, we can only extract fewer patches; we therefore selected more images (5000 images) to build the training and testing datasets. The tile size remains always 128 by 128 pixels with an overlap of 64 pixels.

Figure 5c shows that the learning process appears to converge. Unlike for the Model(Th_{t+f}), both the training and test loss are small.

3.5 Comparing Model(V), Model(Th_{t+f}), and Model(Th_t)

Given the above three models (Model(V), Model(Th_{t+f}), and Model(Th_t), trained on visible images, thermal images with flat and textured areas, and thermal images with flat areas removed, respectively), we carried out a side-by-side comparison using test images from visible and thermal images.

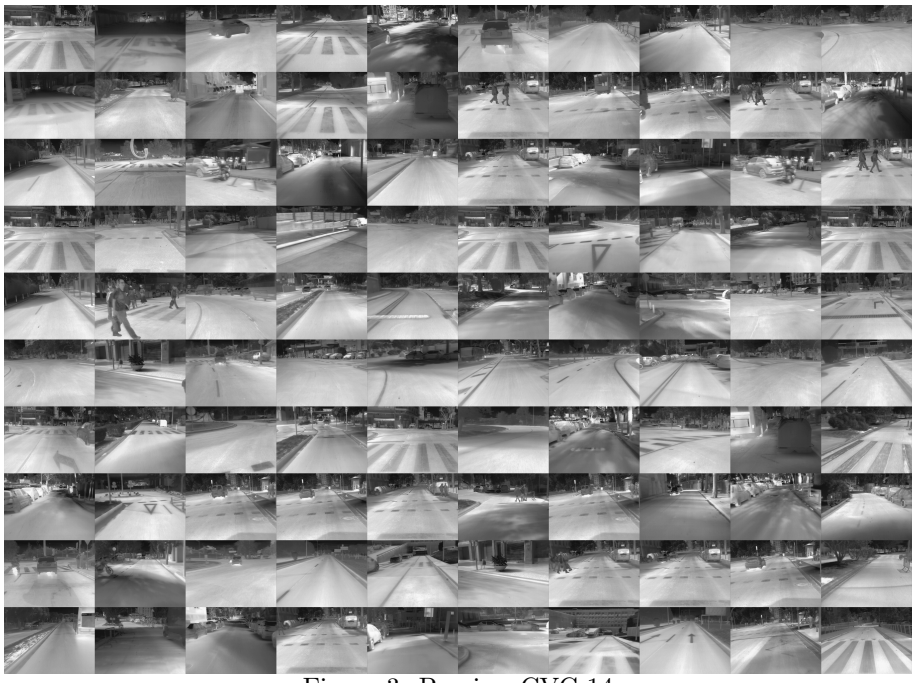
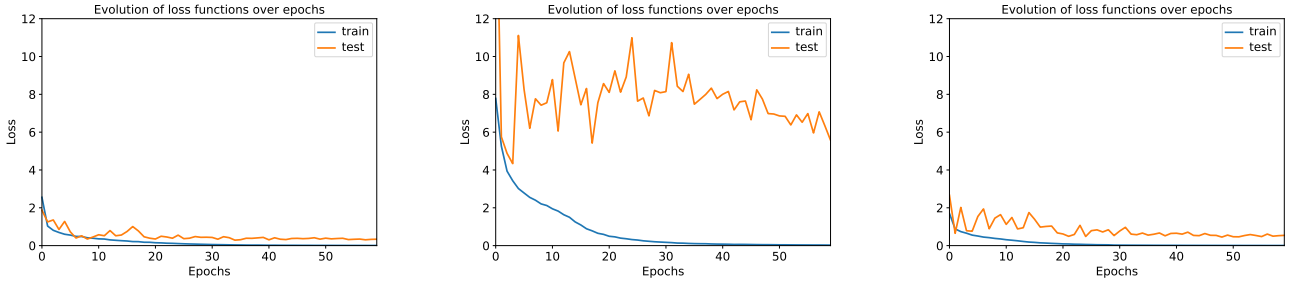


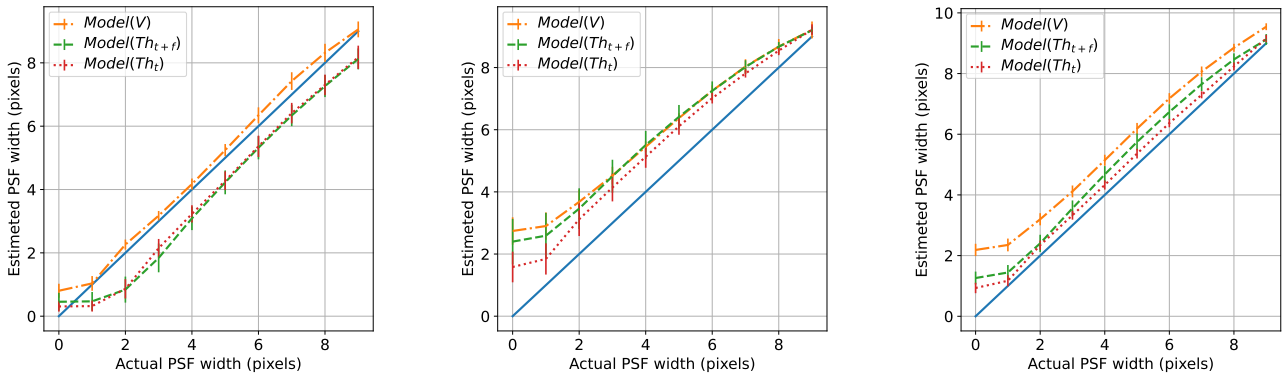
Figure 3: Preview CVC-14

Figure 4: Lack of information on the lower third part of the thermal images from the CVC-14 database.



(a) V (b) Th_{t+f} (c) Th_t

Figure 5: Evolution of loss functions over epochs for the visible (V), thermal with textured and flat areas (Th_{t+f}) and thermal with texture only (Th_t) dataset.



(a) Test data: V (b) Test data: Th_{t+f} (c) Test data: Th_t

Figure 6: Average of the local PSF parameters estimate over 10 images as a function of target parameters, for visible (V), thermal with textured and flat areas (Th_{t+f}) and thermal textured only (Th_t) data set images. The three models (trained on a visible dataset (Model(V)), trained on thermal images with textured and flat areas (Model(Th_{t+f})) and trained on textured thermal images only (Model(Th_t)) were tested with each of the test data sets.

The results are gathered in Figure 6. Figure 6a shows the comparison of the models for V test set (visible images). We can observe that the Model(Th_{t+f}) and the Model(Th_t), which were trained on thermal images, underestimate the PSF parameter, when applied to visible images. Figure 6b shows the performance of the three models for the thermal images data set Th_{t+f}. Model(Th_{t+f}) performs similarly Model(V), while Model(Th_t) gives better estimates for smaller PSF sizes. Finally, Figure 6c compares models on the thermal images with the lower third removed (Th_t). The model trained on visible images, Model(V), tends to overestimate the blurring parameter, while both Model(Th_{t+f}) and Model(Th_t) giving estimates with sub-pixel accuracy, the latter being the most accurate.

We next expanded the local performance evaluations (Section 3.3) and estimated the PSF width with Model(Th_t) (third column in Fig. 7). Model(Th_t) gives better results than Model(Th_{t+f}). Nevertheless, in flat areas it tends to over-estimate the PSF parameter.

4. DISCUSSION

Previous studies have shown that for best performance, it is important to carefully choose a training set that matches the type of images in the target application.[2] Our results confirm these finding, although our use of thermal images also revealed additional caveats. Model(V) gives a good estimate on visible images while both Model(Th_{t+f}) and Model(Th_t) underestimate the width of the PSF when applied to visible images. We observe a strong overestimation of the blur width on Th_{t+f} images, whatever model is used. This can be explained by the fact that the original images are naturally blurred and that the method estimates the total blur, and not only the "added" blur. Indeed, chaining multiple bluring operations is optically and mathematically equivalent to applying a PSF of width equal to the sum of the PSFs at each step. Therefore, if the original image is already blurred (e.g., because the foreground is out of the depth of field range or contains only low frequency features), a synthetically blurred image will show both the lack of sharpness of the original image and the additional blur, resulting in an overestimation of the actual sharpness blur. For Th_t images, Model(V) gives a strong overestimate of the blur width. This may be due to the fact that images in Places205 have a higher resolution than the texture-only thermal images that we considered. The Th_{t+f} model gives a small overestimation of the PSF parameter when applied to the texture-only thermal images Th_t, which could be due to a bias during training/estimation caused by the flat areas of the training data set. This cause is confirmed by the improved performance of the model trained on thermal images with the flat areas removed. A further reason for the better performance, is that the majority of images in the training dataset are of higher contrast and do not show natural blur. Finally, we do not observe a significant difference in terms of learning time or convergence when using

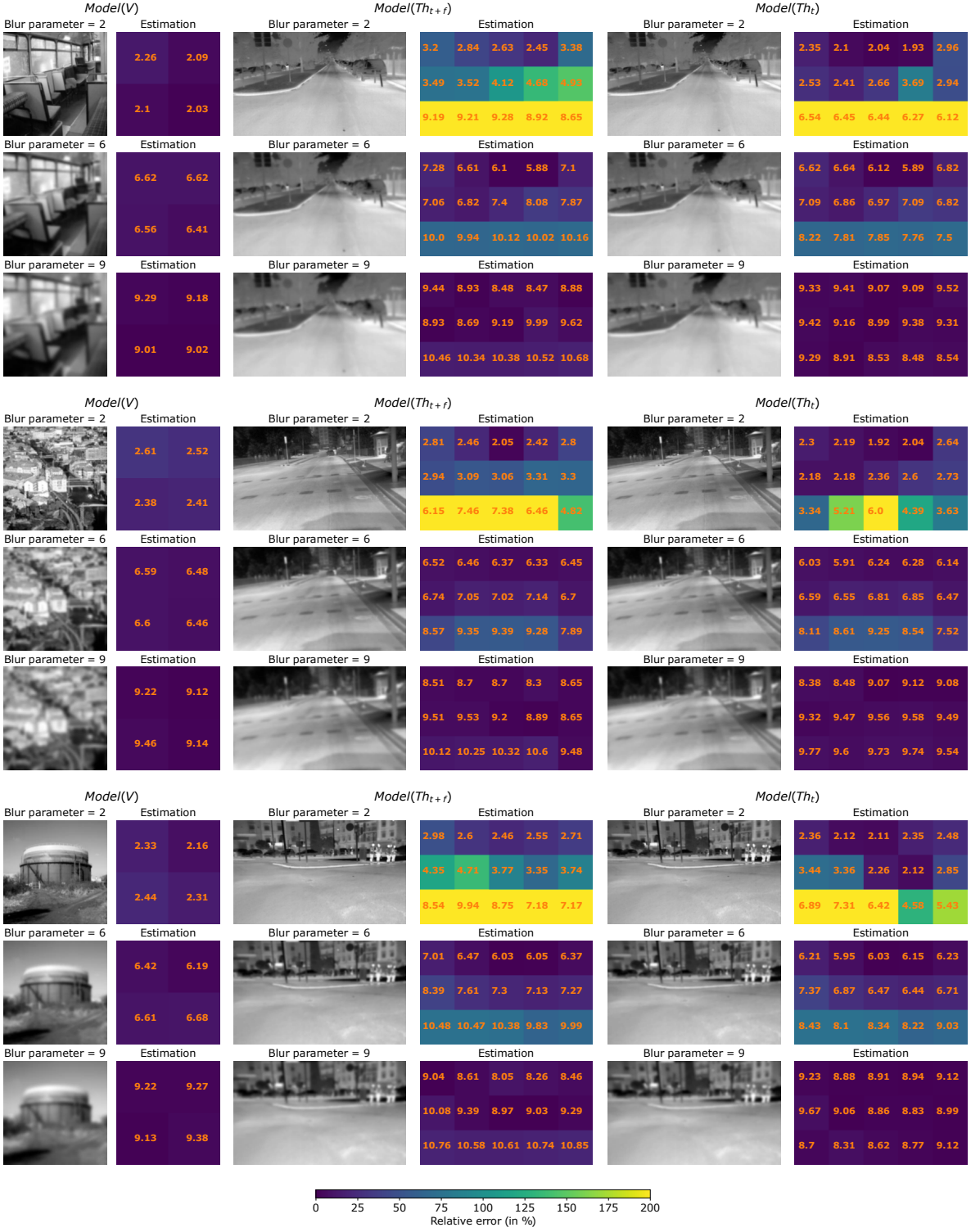


Figure 7: Comparison of the local estimation of the PSF parameter, in areas of size 128 by 128 pixels, for Model(V), Model(Th$_{t+f}$), Model(Th$_t$). Three example images are shown, each blurred with a Gaussian of FWMH equal to 2, 6 or 9. To the right of each blurred image, the map of estimated parameters is shown, where the numbers correspond to the estimated parameter in each region and the background colour represents relative error in that area.

visible or thermal images, as long as the images are sufficiently textured images (slower convergence and poor test loss for Model(Th_{t+f})).

5. CONCLUSION

We have considered the problem of locally estimating the point spread function in thermal images. To this end, we succesfully adapted a local PSF estimation method initially designed for visible light microscopy for the application in thermal imaging.

Our evaluation experiments have highlighted the importance of matching the modality of the training set with that of the end use application, that is, training the model on thermal images. In addition, since our method relies on the availability of high-quality images that are synthetically blured, careful selection of images (possibly leaving out image regions with low contrast) is necessary for accurate estimation. Since the resolution of thermal image databases is typically lower than that of visible light image databases, which are also well curated (balanced picture composition, contrast, high resolution) the achievable accuracy is lower in the thermal case. Our results suggest that curating thermal image datasets of equivalent quality as existing visible light images could further benefit the method.

In conclusion, local PSF parameter estimation via a trained CNN is achievable for thermal image. It requires selecting a large and textured training data set of thermal images. In the future, local deconvolution methods for thermal images could benefit from our proposed PSF estimation method.

REFERENCES

[1] Griffa, A., Garin, N., and Sage, D., "Comparison of deconvolution software in 3D microscopy: A user point of view—part 1," **12**(1), 43–45 (2010).

[2] Shajkofci, A. and Liebling, M., "Spatially-variant CNN-based point spread function estimation for blind deconvolution and depth estimation in optical microscopy," *IEEE Trans. Image Proces.* **29**, 5848–5861 (2020).

[3] Zhu, X., Cohen, S., Schiller, S., and Milanfar, P., "Estimating spatially varying defocus blur from a single image," **22**(12), 4879–4891 (2013).

[4] Sun, J., Cao, W., Xu, Z., and Ponce, J., "Learning a Convolutional Neural Network for non-uniform motion blur removal," in [*IEEE CVPR 2015*],

[5] Gong, D., Yang, J., Liu, L., Zhang, Y., Reid, I., Shen, C., Hengel, A. V. D., and Shi, Q., "From Motion Blur to Motion Flow: A Deep Learning Solution for Removing Heterogeneous Motion Blur," in [*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 3806–3815, IEEE.

[6] Nah, S., Kim, T. H., and Lee, K. M., "Deep multi-scale convolutional neural network for dynamic scene deblurring," in [*IEEE CVPR*],

[7] MIT Computer Science and Artificial Intelligence Laboratory, "Places205 dataset," (2015).

[8] ADAS-CVC, "CVC-14: Visible-FIR day-night pedestrian sequence dataset," (2016).