A DEEP NEURAL NETWORK-BASED METHOD FOR THE DETECTION AND ACCURATE THERMOGRAPHY STATISTICS ESTIMATION OF AERIALLY SURVEYED STRUCTURES

Giorgos Sfikas
Centre for Research & Technology, Information Technologies Institute, Thessaloniki, Greece and Computer Science & Engineering Dpt., University of Ioannina, Ioannina, Greece

João Patacas and Charalampos Psarros
School of Science, Engineering and Design, Teesside University, Middlesbrough, UK

Antigoni Noula, Dimosthenis Ioannidis and Dimitrios Tzovaras
Centre for Research & Technology, Information Technologies Institute, Thessaloniki, Greece

ABSTRACT: *Building thermal output determination is fundamental for the development of energy use optimisation strategies, and can provide important inputs for the estimation of flexibility in demand response strategies. Current building energy assessment procedures are based on design values, not taking into account the uncertainties introduced during the construction and installation processes. The analysis of thermal images of buildings provide the opportunity to carry out the estimation of energy demand based on the actual building performance.*

This research presents a deep neural network-based method for the accurate estimation of thermography statistics from pairs of RGB and thermal images. The proposed method identifies a region of interest (ROI), which is assumed to be a building found approximately at the centre of the image field of view ('target building'). The visible spectrum/RGB input is used to determine the position and outline of the target building in the field of view, and create a pixel-level binary mask with non-zero mask elements corresponding to the target. The binary mask is used to produce an intensity matrix containing only the values that correspond to the building / ROI, and applied to its corresponding thermal image pair. This enables the consideration of the thermal output of the region of interest, as opposed to the whole image, improving the accuracy of the estimation of the ROI's thermography statistics.

KEYWORDS: *digital image processing, automatic structure detection, infrared image processing, deep neural network features, thermal statistics, unmanned aerial vehicle, demand response potential estimation*

1. INTRODUCTION

There is an increasing need for accurate building thermal output determination methods that can be used to assess the actual energy performance of buildings. Currently, the determination of the thermal output of a building presents a series of challenges, due to the difficulty in accounting for the differences between the building design values that are considered in the estimation of energy demand, and the actual performance of buildings. This is largely due to the uncertainties associated with the accuracy of building project documentation, the quality of building materials, and the quality of the construction and installation process.

The estimation of thermography statistics and their usage as a basis for the estimation of energy demand in buildings has been considered as an alternative to current energy assessment methods, which are based on design values (Fokaides et al. 2011; González-Aguilera et al. 2013; Ham & Golparvar-Fard 2012), such as SAP in the UK (BEIS 2014). Accurate methods for the estimation of energy demand in buildings are a key input for the optimization of building energy use, including effective demand response (DR) programs.

Demand response programs are mechanisms developed to ensure that the electricity grid remains stable during times of peak demand (Rodriguez-Trejo et al. 2018). DR programs prompt actions that alleviate the load on the

electricity grid by leveraging the flexibility that users have in their electricity consumption at specific times of the day (Crosbie et al. 2017). Flexibility is established in contracts between companies acting as aggregators and the Transmission Network Operator (TNO) or Distribution Network Operator (DNO). Aggregators acquire flexibility from users (mainly industry and large energy consumers) managing the available assets in response to the grid's requests to increase or reduce electricity consumption or generation (Rodriguez-Trejo et al. 2018; Sisinni et al. 2017). The potential impact of smaller energy consumers on DR flexibility is increasingly being recognized, and challenges such as the lack of integrated tools for optimization, planning and control/management of supply side equipment have been the focus of research efforts such as the Demand Response in Blocks of Buildings (DR-BOB) project (Rodriguez-Trejo et al. 2018).

This research investigates the use of pairs of thermal/infrared (IR) and visible-spectrum (RGB) digital images captured using an UAV to produce features that could aid in identifying the demand response potential of building assets. One such feature that could be correlated to DR potential, is an image of the thermographic characteristics of a building. In the current paper, a novel deep neural network-based method is proposed for the localization and estimation of the thermal output of a region of interest given a pair of thermal (IR) and visible-spectrum (RGB) images. The need to consider pairs of thermal (IR) and visible-spectrum (RGB) images, as opposed to the use of IR images solely, stems from the fact that while IR images already contain temperature data that is used to determine areas of low or high thermal output, they are not appropriate to be used to semantically differentiate objects.

The proposed image processing pipeline has been successfully applied to RGB/thermal image pairs of an existing building captured using a UAV. Preliminary results obtained from the application of the proposed method are expected to contribute to more accurate estimation of baseline and DR flexibility in order to increase the exploitation potential of building assets in DR programs. In addition, the obtained thermal statistics can in perspective be analysed to provide detailed insights in users/customers behaviour. Our proposed method can be leveraged by both smaller energy consumers and energy providers in order to increase the amount of flexibility in DR programs.

The remainder of this paper is structured as follows. In section 2, we present the proposed method, that involves using a pre-trained neural network and unsupervised machine learning methods to localize the structure of interest. In section 3, we discuss the details of image acquisition as well as the test site used for this work, and present qualitative and numerical experiments. We close the paper with conclusions and a discussion of future work in section 4.

2. PROPOSED METHOD

The proposed method involves processing an input of optical and thermal images that depict a building, with an aim to obtain a measurement of thermal characteristics; we shall further assume that this statistic is desirable to be obtained in the form of a thermal intensity histogram (Klette 2014). In order to achieve this aim, as an initial step we use the optical input in order to localize the structure of interest position, and subsequently we use the calculated mask over the thermal input to produce the desired thermal intensity histogram. An overview of the proposed pipeline is provided in Fig. 1.

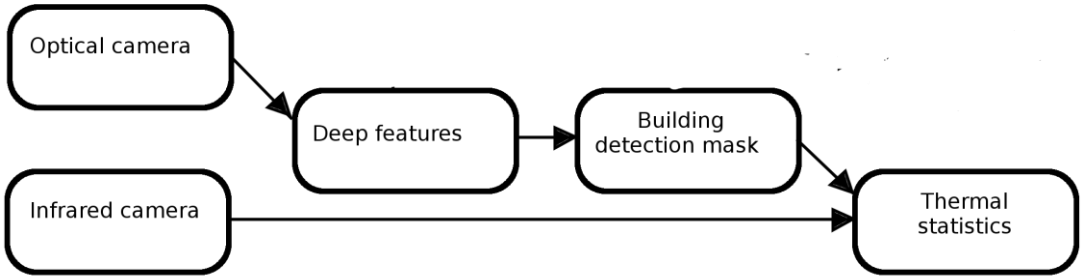

Fig. 1: Proposed image processing pipeline.

2.1 Deep features

The first step of the proposed algorithm involves processing the optical input in order to extract deep features (Sfikas et al. 2016, Retsinas et al. 2019). Deep features are defined as pixel-level cues that are obtained by feeding a pretrained neural network with a particular input – in our case, the input optical image – and calculating the activations of an *intermediate* network layer by performing a standard feed-forward pass (Strang, 2019). The neural network involved is thus used as a tool for feature extraction, instead of the goal it has originally been trained for.

Fig. 2: Typical architecture of feed-forward neural network. Information flows from the input (left) towards the output (right). Intermediate layers correspond to feature maps, each produced by some linear combination of the previous layer or layers (e.g. a convolution). Deep features extraction amounts to simply computing any one of the intermediate feature maps instead of the final network output.

The rationale behind using intermediate layer activations of neural networks as features comes is that we can see a neural network as a representation learning machine (Goodfellow 2016). For example, considering a neural net classifier we can express its output as a linear combination of the penultimate layer activations (composed with a softmax function to produce probability vector). Other, non-neural network classifiers such as the Support Vector Machine (SVM) are based on the same principle, that is perform a linear combination over some set of features and determine whether the output falls in this or the other part of some hyperplane. The difference is that while neural networks follow the same principles as other learning machines, the representation is *learned* during training instead of being hard-coded in the form of 'hand-crafted' features. In the current work, while there is no proper 'training' phase for our method, we assume that the employed networks use weights that have been produced as the optimum of a previous training process, on a third-party dataset, unrelated to the current task (hence the term "pre-training").

In the current application, we have used the Deeplab v3+ neural network, a network that has been proposed for semantic segmentation (Chen et al. 2016). Deeplab comprises a series of convolutional layers, topped by non-linear activations, which eventually lead to a pixel-level set of $K+1$ softmax outputs. The number of K possible outputs is the number of object classes that the network is trained with, plus one reserved for background. The deep feature cues are returned as a matrix of resolution $H \times W$, with each of the cues comprising in general D channels; in other words, on each pixel of the feature map grid, a vector in R^D is produced.

2.2 Calculating the building detection mask

Deep features are processed in order to produce structure of interest segmentations. We perform clustering over the computed deep features, reduced with Principal Component Analysis (PCA) to 8 dimensions. *k-means* is

used to cluster the 8-dimensional features, and k is heuristically set to *3*, initialized with the k-means++ scheme. Between the computed clusters, we have in practice observed that one of the clusters will correspond to the structure of interest, provided of course that the structure of interest covers a significant part of the field of view. We tag the cluster that is situated the closest to the center of the field of view to be the structure of interest cluster. Formally, we use the cluster *j* that minimizes the heuristic:

$$\sum_{n=1}^{N_j} ||x_{jn} - c||$$

where x_{jn} is the position of the n^{th} datum of the j^{th} cluster, and *c* is the position of the field of view center on the image.

2.3 Calculating thermal statistics

After having computed the building segmentation in the previous step, we proceed to calculate thermal statistics over the structure of interest. As the two inputs (optical, thermal) are products of two cameras with different characteristics, both intrinsic (lens, field of view) and extrinsic (different position and pose), the mask cannot be directly applied on the thermal input. The deep feature map, and subsequently the building segmentation map is obtained as a product of the optical cue, hence a step of registration must be applied on the thermal cue. In this work, we register the images manually, by applying an affine transform on one of the inputs so that the two modalities visually match. While this approach will lead to small errors due to misalignment, we must note that alignment of images of different modalities can be a non-trivial task, as different modalities are related to different gradient information, leading to difficulties in feature-based matching that typically uses gradient information to work (Klette 2014).

After applying the segmentation mask over the thermal input, we obtain the set of thermal intensities that correspond to the structure of interest surface. These can be in turn used to compute thermal intensity statistics. We have computed histograms of thermal intensities, as well as computed thermal intensities means, though in practice any other, possibly more complex, statistic can be computed over the features. The advantage that stems from the proposed scheme, is that *only the actually relevant* thermal intensity values are obtained, thus any statistic over thermal values will objectively be more accurate. In the numerical results section, we show that indeed in practice a great disparity on the thermal estimate using our method versus not using it can exist, validating the usefulness of the method.

3. EXPERIMENTAL RESULTS

3.1 Experiment site

The site used for the aerial image acquisition is the Smart Home building, located at CERTH's facilities in Thessaloniki, Greece. It is a rapid prototyping & novel technologies demonstration infrastructure resembling a real domestic building where occupants can experience actual living scenarios. It was chosen because it is equipped with features found at a typical modern domestic building (PV array, solar water heating, heat pump units) and is located within adequate distance from other structures, making it convenient to safely fly a drone for aerial surveys.

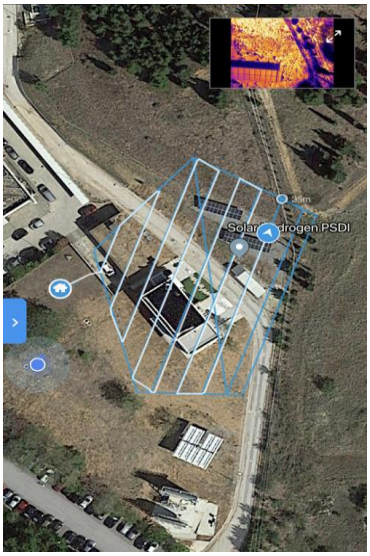

Fig. 3: UAV typical flight plan used for our experiment over the test site.

3.2 Image acquisition

The UAV used was a DJI Matrice M200, equipped with a DJI Zenmuse XT2 Visual+Thermal Imaging camera. The visual camera uses a 4K (Ultra HD) sensor with a resolution of 3840x2160 pixels while the thermal camera uses a Vanadium Oxide Microbolometer with an output resolution of 640x512 pixels. Video capture of the cameras were at 29.97 frames per second for the visual imager and 30Hz for the thermal imager. A flight plan (Fig. 3) was programmed for the drone to perform an aerial survey as an autonomous mission, followed by a manual flight around the building for video acquisition. In Fig. 4 an example of an input IR/RGB image pair is provided.

Fig. 4: Example input pair. An optical camera frame (left) and corresponding thermal input frame (right). Images were captured with cameras mounted on an Unmanned Aerial Vehicle (UAV), surveying the structure of interest.

3.3 Results

Let us note that the proposed pipeline is completely unsupervised, in the sense that it requires no annotated data or a training phase to run. We have run tests using either ADE20k or the Pascal VOC pretrained weights[1]. These have been trained on the respective homonymous datasets, which comprise 150 and 20 classes respectively. We have found that either of the two weight sets produces useful results, segmenting the target structure correctly; this is certainly a useful feature, especially for the Pascal VOC weights, as training in this dataset has been done in a set that does not contain buildings or structures as a separate class.

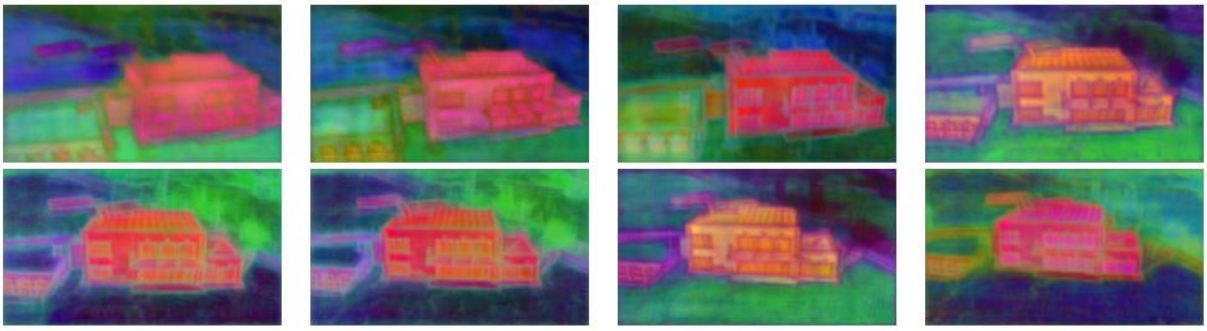

Fig. 5: Visualization of deep features for input images. High-dimensional, pixel-wise deep features were reduced to 3 dimensions with PCA and shown here as pseudo-colorized images. Note that on each frame semantically similar objects correspond to similar colors.

Furthermore, with preliminary tests on buildings other than the location and structure discussed in subsection 3.1, we have noted that the proposed structure estimation algorithm runs quite well in a wide range of structure types. Regarding computation of deep features, we have used the ReLU activation (Strang, 2019) of the last feature map before the network output ("*decoder/decoder_conv1_pointwise*"). Deep features originally are of dimension equal to 256, i.e. each pixel-level cue is a vector in the R^{256} space. These vectors are reduced to the R^3 space using Principal Component Analysis (PCA) (Strang, 2019), solely for the purposes of visualization. Each of the three obtained channels is then assigned to one of the optical Red, Green or Blue channels, and after scaling values to an 8-bit range (0..255) the visualizations shown in Fig. 5 are produced. Note that the visualized features do indeed carry semantic information, as areas of similar color correspond to *semantically* same or similar areas. For example, the whole are of the building is marked with a similar reddish color, even though its image in either the optical or thermal domain is visibly diverse, in the sense that it comprises areas with different colors and/or different thermal intensities (see Fig. 5).

Qualitative results can be examined in the subsequent Figures 6 and 7, where the building detection mask step is shown for the test frames (Fig. 6), and the final thermal statistic calculation step is performed (Fig. 7).

1 Publicly available under https://github.com/tensorflow/models .

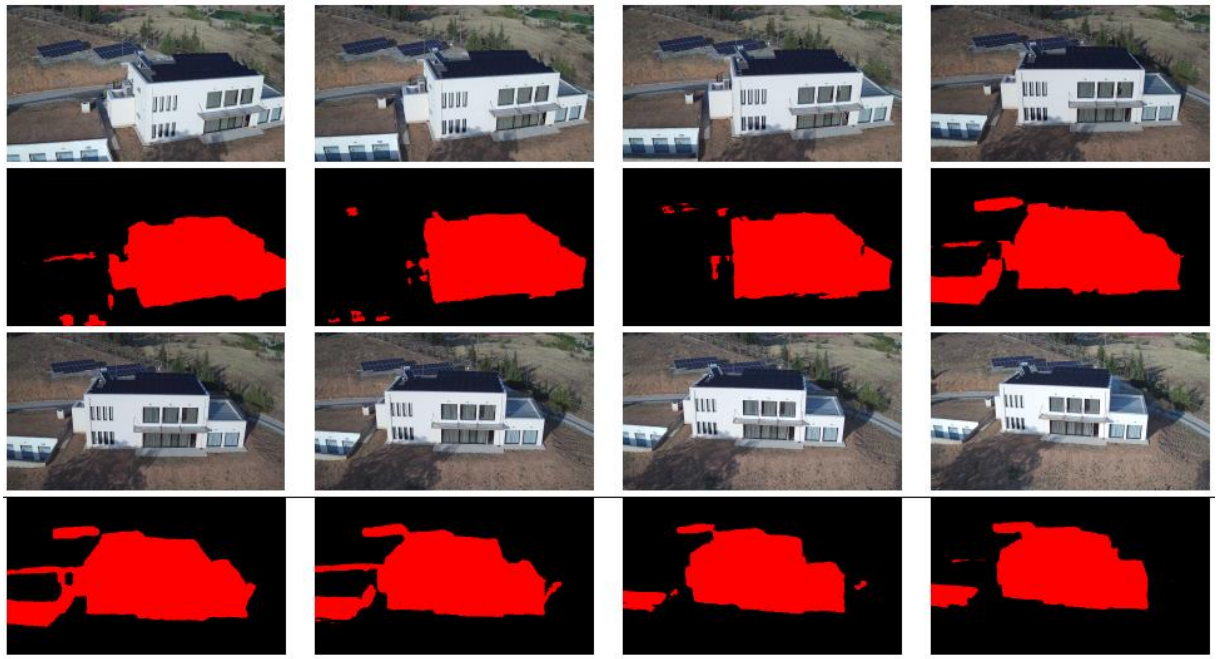

Fig. 6: Detection of main structure in each frame (2nd and 4th row) juxtaposed to original optical input frames (1st and 3rd rows). The areas marked in red correspond to the automatically detected structure.

We report numerical results on Table 1. The reported figures are disparity metrics computed over a number of frames of the captured footage, and they serve to measure the divergence between the statistics of the infrared / thermal image as computed when taking into account the proposed structure localization scheme versus to not taking it into account and just computing statistics over the whole infrared frame each time. In particular, we compute disparity using the following formula:

$$disparity = \frac{|\mu(IR_{full}) - \mu(IR_{proposed})|}{\mu(IR_{full})}$$

where μ(.) denotes mean values over either the while infrared input frame (IR_{full}) or the infrared frame masked using the localization result with the proposed scheme ($IR_{proposed}$). The reported figures show that the disparity between the two cases is certainly not at all negligible, with figures ranging around the 15% mark. In practice, this means that a 'naïve' estimate of the structure thermal signature, i.e. without localizing it correctly first, will lead to a significant error in estimating the correct temperature of the region of interest. Note also from the shown infrared intensity histograms (Fig. 7), that not only the means and peaks of the two infrared intensity distributions differ significantly, but also the whole distributions as a whole. The structure thermal distribution is characterized by a second, low intensity peak, which corresponds to the photovoltaic units found on the structure roof (shown in histogram figures in blue color). This is completely missed by the naïve thermal estimate (shown in histogram figures in green color), which erroneously shows a roughly unimodal thermal distribution. Aside from being an important aspect in the context of computing a thermography statistic estimate *per se*, using the bimodal nature of the histogram could have implications in further digital image processing steps, as for example performing a simple intensity histogram-based algorithm to localize the low-intensity cluster that corresponds to the photovoltaic units.

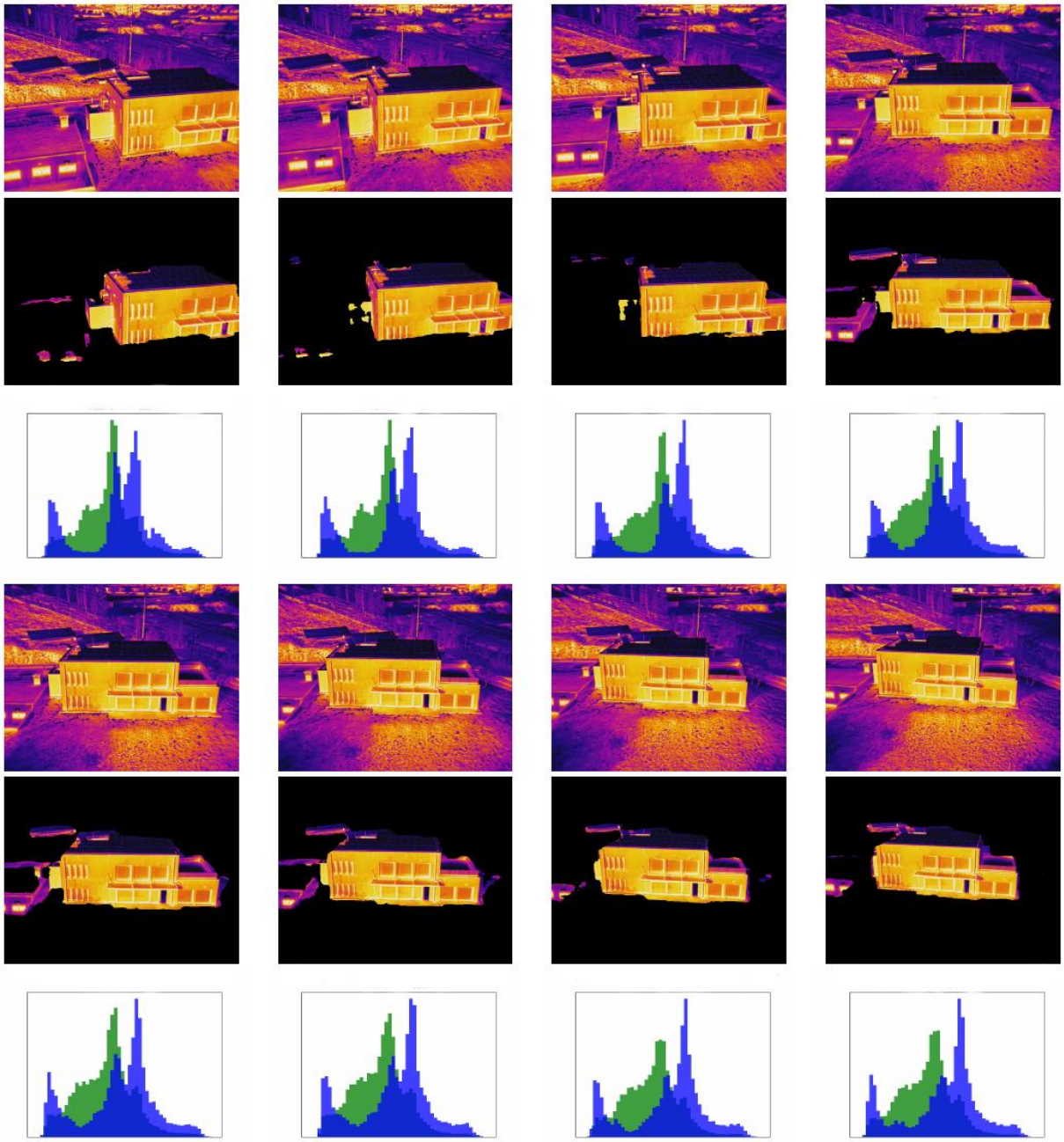

Fig. 7: Segmentation of thermal imaging intensities that correspond to the structure of interest (2nd and 4th rows), juxtaposed to full thermography scans (1st and 3rd rows). Thermography statistics using our method versus using all thermography intensities, for each frame are shown (3rd and 6th row). In all cases there is a clear disparity between thermal statistics versus only the structure intensities, justifying the use of our detection method. In the histograms, the horizontal axis comprises bins from minimum to maximum thermal intensity, and the vertical axis corresponds to the number of pixels with the given thermal intensity.

Table 1: Disparity between thermal statistics computed using the proposed method versus computing whole thermal image frames. Figures are computed over the 8 test images shown in figures 2-4. In all cases, disparity is far from being negligible; hence, using the proposed method is necessary to produce accurate thermal statistics.

Percentage offset								
Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Mean +- St.dev.
12.9%	12.8%	12.5%	14.6%	14.9%	15.8%	18.2%	17.6%	14.9% +- 2%

4. CONCLUSION AND FUTURE WORK

In this paper we have presented a method that uses a dual input of optical (RGB) and thermal inputs in order to localize a structure of interest and simultaneously obtain an accurate reading of its thermal characteristics. The proposed method is completely unsupervised, with no training or manual annotation of the structure of interest required. A neural network is used to produce deep features, that are processed to produce the required outputs. The pretraining of the network is not even required to have 'seen' a building class, and we have checked that valid outputs are obtained with two different sets of network pretrained weights. In perspective, the method can be extended to perform detection of other assets, like photovoltaic units. Also as future work, we plan to use multispectral image processing methods (Sfikas et al. 2011) to align and process the optical and thermal pairs more accurately, or integrate the proposed method with an Internet-of-Things (IoT)-based scheme (Sfikas et al. 2016). Furthermore, we look forward to applying thermography-based image processing to other tasks related to construction and materials (for example, relation of temperature and concrete cracking risk) (Kanavaris et al. 2017, Kanavaris et al. 2019) and exploring other uses of UAV-obtained thermal statistics of surveyed structures.

ACKNOWLEDGEMENTS

The work presented was carried out as part of the eDREAM project which is co-funded by the EU's Horizon 2020 framework programme for research and innovation under grant agreement No 774478. The authors wish to acknowledge the European Commission for their support, the efforts of the project partners, and the contributions of all those involved in eDREAM.

REFERENCES

BEIS (2014) *Standard Assessment Procedure Guidance on how buildings will be SAP energy assessed under the Green Deal and on recent changes to incentivise low carbon developments*. Available at: https://www.gov.uk/guidance/standard-assessment-procedure Accessed on: 16th May 2019

Bishop C.M. (2006), Pattern Recognition and Machine Learning, Springer.

Chen L.C., Zhu Y., Papandreou G., Schroff F., Hartwig A. (2018), Encoder-Decoder with Atrous Separable convolution for semantic image segmentation, the European Conference on Computer Vision (ECCV), pp. 801-818

Crosbie T., Vukovic V., Short M., Dawood N., Charlesworth R., Brodrick P. (2017). Future Demand Response Services for Blocks of Buildings. *Smart Grid Inspired Future Technologies* 10.1007/978-3-319-47729-9_13.

Fokaides Paris A., Soteris A. Kalogirou (2011), Application of infrared thermography for the determination of the overall heat transfer coefficient (U-Value) in building envelopes, *Applied Energy*, Volume 88, Issue 12, 2011, Pages 4358-4365, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2011.05.014.

González-Aguilera D., S. Lagüela, P. Rodríguez-Gonzálvez, D. Hernández-López (2013), Image-based thermographic modeling for assessing energy efficiency of buildings façades, *Energy and Buildings*, Volume 65, 2013, Pages 29-36, ISSN 0378-7788, https://doi.org/10.1016/j.enbuild.2013.05.040.

Goodfellow I., Bengio Y., Courville A. (2016), "Deep Learning", MIT Press

Ham Y. & M. Golparvar-Fard (2012) Rapid 3D Energy Modeling for Retrofit Analysis of Existing Buildings Using Thermal and Digital Imagery. *LiDAR Magazine*, Vol. 2, No. 4

Kanavaris F., Soutsos M., Chen J.F. (2017), Effect of temperature on the cracking risk of concretes containing ground granulated blast-furnace slag, Second International RILEM/COST Conference On Early Age Cracking and Serviceability in Cement-based Materials and Structures EAC02, At Brussels, Belgium

Kanavaris F., Kaethner S. (2019), Ciria guide C766: An overview of the updated Ciria C660 guidance on control of cracking in reinforced concrete structures, International Conference on Sustainable Materials, Systems and Structures

Klette R. (2014), "Concise Computer Vision", Springer, London

Retsinas G., Louloudis G., Stamatopoulos N., Sfikas G., Gatos B., An Alternative Deep Feature Approach to Line Level Keyword Spotting (2019), IEEE Computer Vision and Pattern recognition (CVPR), Long Beach (CA), USA

Rodriguez-Trejo S., Crosbie T., Dawood M., Short M., Dawood N. (2018) From Technology Readiness Levels (TRL) to demonstrating Demand Response in Blocks of Buildings: Upgrading technical and social systems, *Proceedings of the 17th International Conference on Computing in Civil and Building Engineering (ICCCBE) 2018*, June 5 7, 2018, Tampere, Finland

Sisinni M., Noris F., Smit S., Messervey T., Crosbie T., Breukers S. and Van Summeren L. (2017) Identification of Value Proposition and Development of Innovative Business Models for Demand Response Products and Services Enabled by the DR-BOB Solution, *Buildings*, Special Issue Selected Papers from Sustainable Places 2017 (SP2017) Conference.

Sfikas G., Akasiadis C., Spyrou E. (2016) Creating a smart room using an IoT approach, Proceedings of the Workshop on AI and IoT (AI-IoT), 9th Hellenic Conference on Artificial Intelligence, Thessaloniki, Greece

Sfikas G., Heinrich C., Zallat J., Nikou C., Galatsanos N. (2011), Recovery of polarimetric Stokes images by spatial mixture models, Journal of the Optical Society of America (JOSA A)

Sfikas G., Patacas J., Noula A., Ioannidis D., Tzovaras D. (2019), "Building thermal output determination using visible spectrum and infrared inputs", International Conference on Energy and Sustainable Futures (ICESF), Nottingham, United Kingdom

Sfikas G., Retsinas G., Gatos B., Zoning Aggregated Hypercolumns for Keyword Spotting (2016), International Conference on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China

Strang G., "Linear algebra and Learning from data" (2019), Wellesley-Cambridge Press