Watching the Air Rise: Learning-Based Single-Frame Schlieren Detection

Florian Achermann[1], Julian Andreas Haug[1], Tobias Zumsteg[1], Nicholas Lawrance[1,2], Jen Jen Chung[1,3], Andrey Kolobov[4] and Roland Siegwart[1]

***Abstract*— Detecting air flows caused by phenomena such as heat convection is valuable in multiple scenarios, including leak identification and locating thermal updrafts for extending UAV flight duration. Unfortunately, the heat signature of these flows is often too subtle to be seen by a thermal camera. While convection also leads to fluctuations in air density and hence causes so-called *schlieren* – intensity and color variations in images – existing techniques such as Background-oriented schlieren (BOS) allow detecting them only against a known background and from a static camera, making these approaches unsuitable for moving vehicles. In this work we demonstrate the feasibility of visualizing air movement by predicting the corresponding schlieren-induced optical flow from a *single* greyscale image captured by a moving camera against an unfamiliar background. We first record and label a set of optical flows in an indoor setup using standard BOS techniques. We then train a convolutional neural network (CNN) by applying the previously collected optical flow distortions to a dataset containing a mixture of real and synthetically generated images to predict the two-dimensional optical flow from a single image. Finally, we evaluate our approach on the task of extracting the optical flow caused by schlieren from both a static and moving camera on previously unseen flow patterns and background images.**

I. INTRODUCTION

Visualizing air flow patterns caused by heating [1], physical absorption [2], chemical reactions [3], or shock waves [4] is valuable in multiple fields such as leak detection, the study of boundary layer detachment, heat transfer, or locating thermal updrafts for extending small uncrewed aerial vehicle (sUAV) flight duration. In the latter case, for example, existing flight controllers allow a fixed-wing sUAV to exploit thermal updrafts for soaring, the way birds [5]–[8] and human glider pilots do, if the sUAV happens to stumble upon such an updraft [9]–[12]. However, actually finding these updrafts presents a problem. Predicting their locations using consistent thermal infrared (TIR)-optical mapping techniques such as MultiPoint [13] based on the ground temperature and color comes with significant uncertainty. Directly observing the thermal columns in the sUAV's vicinity using an onboard sensor could be very helpful. In our work, we take a step towards making this a reality.

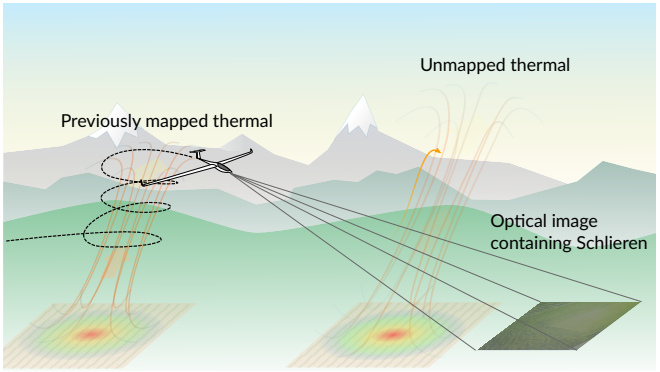

Fig. 1: The difference in the refractive index between the thermal and surrounding air causes schlieren in the optical image. Detecting these patterns enables sUAVs detecting thermals from far away.

Air flows such as thermals give rise to variations in the air's refractive index and lead to subtle bending of light rays according to Snell's Law [14] as the rays traverse through these fluctuations. The resulting brightness and color changes are called *schlieren*, from the German word for 'streaks' [15]. Usually, schlieren are invisible to the human eye unless they are caused by large temperature differences, e.g. immediately above a road surface on a hot sunny day. However, they can be visualized in a controlled lab setup using multiple lenses and a point light source as a shadowgraph measuring the second derivative of the density [16]. BOS methods visualize the schlieren with a more generic setup, only requiring one camera but assuming a known, high-texture background [17], [18]. Optical flow techniques based on intensity variations, between successive frames or compared against an undisturbed reference frame, can then be used to extract the image distortion caused by the schlieren.

In scenarios such as sUAV flight, however, the background is generally not known and BOS methods are not applicable directly: although they have been applied aboard aircraft, they require multiple passes over the recording area to capture the background [19]. Reference-free BOS techniques don't require a reference background and can detect the schlieren using a stereo setup of high-quality cameras, but are computationally complex and require high-texture backgrounds [20]. Since we are interested in simply identifying areas with schlieren and aren't trying to find the true flow, computing the optical flow between two consecutive frames could serve as a viable approximation. However, movement

This work was supported by Microsoft Swiss Joint Research Center under Contract No. 2019-038 "Project Altair: Infrared Vision and AI Decision-Making for Longer Drone Flights".

[1] Autonomous Systems Lab, ETH Zürich, Zürich 8092, Switzerland {acfloria, haugj, tzumsteg, rsiegwart}@ethz.ch

[2] Robotic Perception and Autonomy, CSIRO Data61, QLD 4069, Australia, nicholas.lawrance@csiro.au

[3] School of Electrical Engineering and Computer Science, The University of Queensland, Staff House Road, Brisbane, QLD 4072, Australia, jenjen.chung@uq.edu.au

[4] Microsoft Research, One Microsoft Way, WA 98052, USA, akolobov@microsoft.com

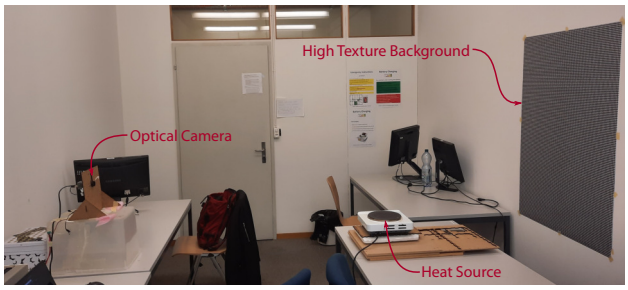

Fig. 2: The indoor setup to record thermally induced flows that comprise the ROSCINE dataset. The known high-texture background provides optimal conditions for computing the optical flows caused by hot air above the heat source. The background is a greyscale image with sinusoidal brightness changes in both directions with a wavelength of $8.7\,\mathrm{mm}$, equivalent to $8\,\mathrm{px}$ in the captured image.

of the sUAV causes viewpoint changes between consecutive frames. Perfect alignment of the two frames is impossible in a three-dimensional environment. As a result, artifacts in the optical flow overshadow the sub-pixel flow of the schlieren.

In this work we propose a method called DYBS (**DY**namic-**B**ackground **S**chlieren) for detecting the optical flow of the schlieren from a **single** optical greyscale image using a CNN. We first collect a dataset of **r**eal **o**bserved **s**chlieren in a **c**ontrolled **in**door **e**nvironment (**ROSCINE**), in a lab setting with a static camera and various high-texture backgrounds and different heat sources. We extract the optical flow due to the schlieren with traditional BOS techniques using the known undistorted background and treat this optical flow as the ground truth. Then we generate a dataset of imagery containing schlieren with the respective ground-truth optical flows. The dataset is composed of real imagery from the indoor setting and synthetically generated images. In the latter case we treat the optical flow of the schlieren as a distortion map that is applied to images from the Places dataset [21] to simulate the appearance of schlieren on different backgrounds and textures. We train the network with a mixture of the real and synthetic images and finally evaluate it on held-back data and real-world imagery.

II. GROUND-TRUTH FLOW GENERATION

We collected the high-quality ROSCINE dataset in an indoor setting optimized for generating different labelled schlieren patterns with multiple high-texture backgrounds. We used a global shutter optical camera (UI-5261SE Rev. 4 with a $16\,\mathrm{mm}$ focal length lens) to capture images at $25\,\mathrm{Hz}$. We used two different heat sources, a larger and a smaller electric heat plate at different temperature settings to generate various shapes of thermal updrafts. The high-texture backgrounds allow for accurate computation of the sub-pixel optical flow [22]. The setup is shown in Fig. 2.

We evaluated different background patterns and optical flow algorithms to optimally capture the schlieren patterns and reduce the measurement noise. Since even in our controlled setup we don't have access to ground truth schlieren we subjectively evaluated the different approaches by their signal to noise ratio and whether they could generate the sub-pixel optical flow of the schlieren. The greyscale backgrounds used were checkerboard patterns of different sizes, Perlin noise [23], sinusoidal patterns, and a pattern consisting of multi-scale black and white squares. In our tests, all selected background patterns were sufficient for computing the small scale optical flows. To compute the optical flow of the schlieren we tested the Farnebäck [24], Horn-Schunck [25], Lucas-Kanade [26], PCAFlow [27], DeepFlow [28], and SPyNet [29] algorithms. SPyNet, PCAFlow and Lucas-Kanade failed to compute the small-scale optical flow, and the magnitude of the noise was higher than the flow due to the schlieren. Farnebäck, DeepFlow, and Horn-Schunck were all able to capture the schlieren. However, DeepFlow produced an overly smooth image and failed to pick up some of the small scale details. Farnebäck exhibited the highest noise levels of these three algorithms. Finally, we selected the Horn-Schunck algorithm to compute the ground-truth flow data for ROSCINE.

We maintained a constant distance between the camera and the background wall but varied the position and size of the heat source to record varying shapes and magnitudes of schlieren. Small vibrations or movements of the camera relative to the background during the recordings introduced small optical flow biases in each direction in the order of $0.0\,\mathrm{px}$ to $0.2\,\mathrm{px}$. Since only a small portion of the recorded image contained schlieren we could determine the bias by computing the median flow value in each direction and subsequently subtract the bias to correct the flow values.

In total we recorded imagery with 11 different background patterns and for each setting used the large heat source in three and the small one in two different heat settings. We also varied the distance between the heat source and the camera, once placing it close to the camera and once closer to the background. In total this results in 110 different setups with observed schlieren of different shapes and magnitudes. A few examples for the different flow shapes are shown in Fig. 3 where the hue and saturation indicate the direction and magnitude of the flow, respectively. We observed that for the same heat source the optical flow magnitudes increase when the heat source is closer to the camera than the background. The flow generated by the large heat plate is almost immediately turbulent, while the flow above the small

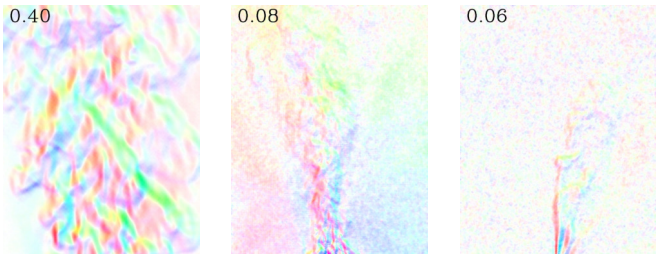

Fig. 3: Examples of different types of ground-truth flows recorded in the indoor setting. The left flow pattern is generated from the large heat plate close to the camera, in the middle with the same heat source far away from the camera. The flow on the right is captured from the small heat plate close to the camera. The maximum flow norm in pixels is shown in the top left corner for each image.

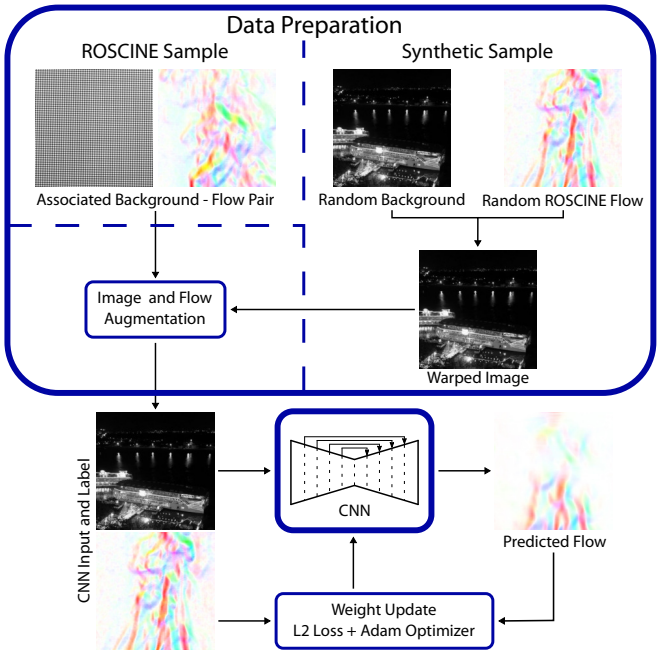

Fig. 4: Pipeline overview to train DYBS with synthetic and real images. The input data to DYBS are either real samples from the ROSCINE dataset or synthetically generated samples with a random background distorted with a ground-truth ROSCINE flow.

heat plate is laminar until approximately the middle of the captured image. The maximum flow magnitude per image varies throughout the dataset between $0.07\,\mathrm{px}$ to $1.08\,\mathrm{px}$.

Certain setups showed strong artifacts in the optical flows computed by the Horn-Schunck algorithm due to movement of the camera, the background or slight illumination changes to the calibration image. These cases were manually detected and the data discarded. The ROSCINE dataset consists of the 81785 $\langle$optical flow, background$\rangle$ pairs, split across the different cases into training and test subsets containing 61879 and 19906 samples, respectively.

III. NEURAL NETWORK TRAINING FOR SINGLE FRAME SCHLIEREN DETECTION

We developed the pipeline to train DYBS to predict the optical flow due to schlieren using only one single grayscale image as input. The full pipeline is displayed in Fig. 4.

Network Architecture. We chose a CNN-based architecture due to their efficiency and ability to predict based on local spatial features. The architecture used in this work is based on the UNet with minor changes [30]. While originally created for pixel-level segmentation, UNet-style models have demonstrated strong performance for precise reconstruction due to the skip connections passing spatial information between down- and up-sampling layers. We have the same depth as the original UNet (four pooling and upsampling layers) and utilize the skip connections but replace the fully connected layers at the bottleneck with convolutions. This change results in a fully convolutional network that can handle inputs of varying sizes above the minimum of $16\,\mathrm{px} \times 16\,\mathrm{px}$. We extend the nonlinearity after each convolution with a BatchNorm layer to stabilize the training by reducing the internal covariance shift [31]. The input to the CNN is a single greyscale image. The pixel values are mapped to $[0, 1]$. The model predicts the two-dimensional optical flow for each input pixel.

Dataset We trained the neural network on the dataset of recorded schlieren flows. We used a combination of samples from the ROSCINE dataset that already contain the schlieren (associated pairs of the recorded high-texture background with schlieren and the corresponding flow from our controlled indoor environment) and synthetically generated samples where the schlieren distortions (ground-truth flows) from ROSCINE applied to existing images from the Places Standard dataset [21]. Using the 1.8 million images from the Places training dataset with widely varying textures and scenes as background prevented the network from overfitting to the high-texture backgrounds used during data collection, which exhibit much less variety with only 11 different patterns. We utilized the high-resolution images from the Places Standard dataset above the minimum size of $480\,\mathrm{px} \times 480\,\mathrm{px}$.

A sample with a background from the Places dataset was constructed in the following way. Since the background images are smaller than the ground-truth flows, we randomly cropped the flows to the same size as the background at $480\,\mathrm{px} \times 480\,\mathrm{px}$. To randomize the flow directions, we flipped the cropped flow in the x- and y-directions with a probability of 0.5. The extracted ground-truth optical flows (Sec. II) are essentially warp maps describing the magnitude and direction in which the schlieren shifted the pixels of the undistorted background. As the final step of generating synthetic samples we apply an extracted ground-truth flow label as a warp to a novel (undistorted) Places background image. The ROSCINE samples already contained the schlieren and did not require preprocessing. Finally we randomized the orientation of the image-flow pair by flipping the flow and the image randomly along each axis with a probability of 0.5.

During training, we randomly selected the sample type. A sample from ROSCINE was chosen with probability p_R, and a sample with a flow from ROSCINE but a background from the Places Dataset was picked with probability $1 - p_R$. This allowed us to balance the training between the real samples with the limited variety background texture but including the camera sensor noise and the synthetic images with highly variable backgrounds.

Loss. We used a regular mean squared error (MSE) loss between the optical flow predicted by the network and the ground-truth flow to train the network.

Learning Framework Setup. The training pipeline was implemented with the PyTorch framework [32] using the Adam optimizer [33] with a batch size of 20 samples to optimize the model weights. We set p_R to 0.5 resulting in the CNN observing an equal amount of synthetic and real samples during training.

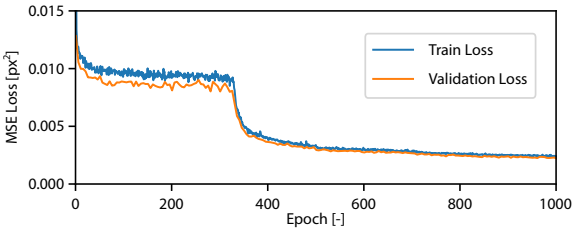

Fig. 5: The training and validation MSE loss during the first stage of the curriculum learning. The training losses were averaged over 10'000 update steps equalling one epoch. The validation loss was computed every fifth epoch.

DYBS Training Schedule. Training DYBS directly on the dataset containing all different flow patterns resulted in strong overfitting of the real samples and no progress on the synthetically generated ones. By training the model with a curriculum learning strategy [34] we were able to mitigate the overfitting and trained DYBS to predict the flow for both background types.

In the first stage of the curriculum, we trained the model only on a limited set of flow patterns from only one background setting with the large heat plate (7231 flows, 11.6% of all recorded flows) for five million update steps with a learning rate of 1×10^{-4} (500 epochs) and another five million steps with a decreased learning rate of 2.5×10^{-5}. The training and validation MSE loss over the 1000 training epochs of this first learning stage are shown in Fig. 5. The training loss was averaged over one epoch composed of 10'000 batches. We computed the validation loss every fifth epoch. After an initial decline, the loss values stagnated until roughly epoch 330. Up to this point the model learned to predict samples from ROSCINE well but failed on the synthetic samples concocted from ROSCINE's flows and the Places Dataset's backgrounds. The second decline in MSE loss shows the model learning to predict the latter samples as well. This phenomenon was consistent across different training runs, though the starting epoch for learning to predict both sample types was slightly different in each run.

In the second curriculum stage, we let the model train for another 5 million update steps with a learning rate of 1×10^{-4} on the full dataset. This forced the model to generalize to different flow shapes and real background images.

IV. EXPERIMENTS

We evaluated DYBS in a series of experiments with increasing difficulty. In the first set of experiments, we used held-out flow data from our ROSCINE dataset, recorded in the indoor setup, together with images from the Places Validation Standard dataset to compare the DYBS predictions to the optical flow from the BOS algorithm. In the second experiment, we tested if the network could generalize to different flow patterns in an indoor setting with a static camera and a high texture background. We generated different flow patterns by deflecting the flow with wind or directly with a plate over the heat source. In the last set of experiments we used recordings from a moving camera in an indoor and outdoor setting to test how well the network can generalize to camera motion, lighting changes and various real backgrounds.

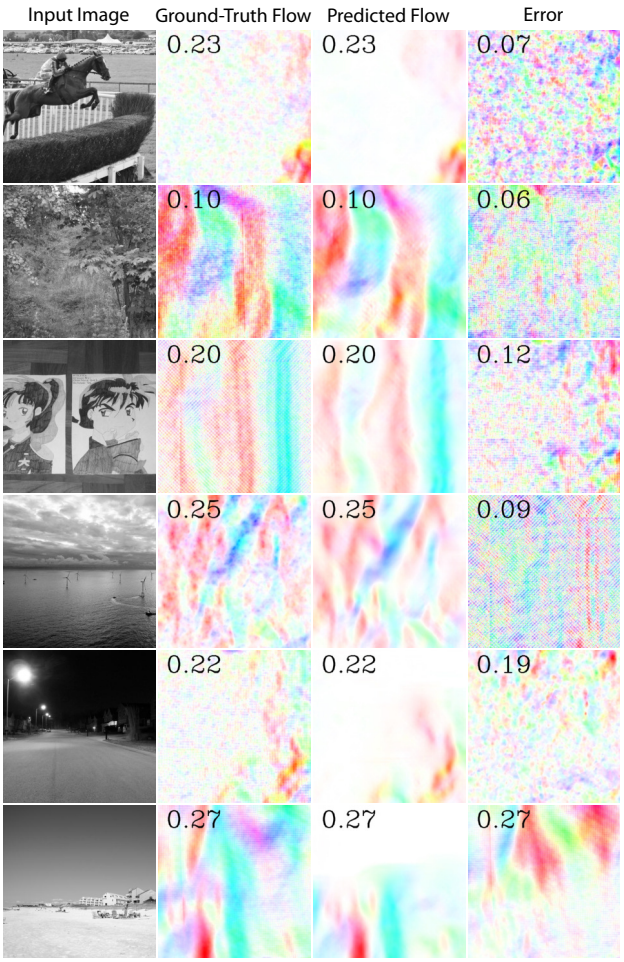

Fig. 6: Qualitative prediction results of DYBS on previously unobserved samples using images from the Places dataset as background (previously unseen images with previously unseen flow distortions applied). The maximum flow magnitude in pixels for each picture is indicated in the top left corner. For clarity, the same magnitude color scale is used for both the predicted and ground-truth flows.

Experiment 1: Test Dataset. We evaluated DYBS on samples with the different background types separately with previously unobserved flow patterns and background images.

DYBS predicts the flow for the sample images from the Places dataset with a mean magnitude prediction error of $0.017\,\mathrm{px}$ at a mean flow magnitude of $0.022\,\mathrm{px}$. At first glance, the prediction error seems relatively high. However, some representative predictions displayed in Fig. 6 show that DYBS detects the sub-pixel flow accurately for different flow shapes and magnitudes as well as highly varying backgrounds. The ground-truth flows from the ROSCINE dataset contain small-magnitude artifacts corresponding to the background patterns, as evident from the top four rows in Fig. 6. The DYBS predictions do not contain such artifacts and thus are much smoother than the ground-truth flows, contributing to a higher error metric despite modeling the real flows well. In case of no texture, such as the sky in

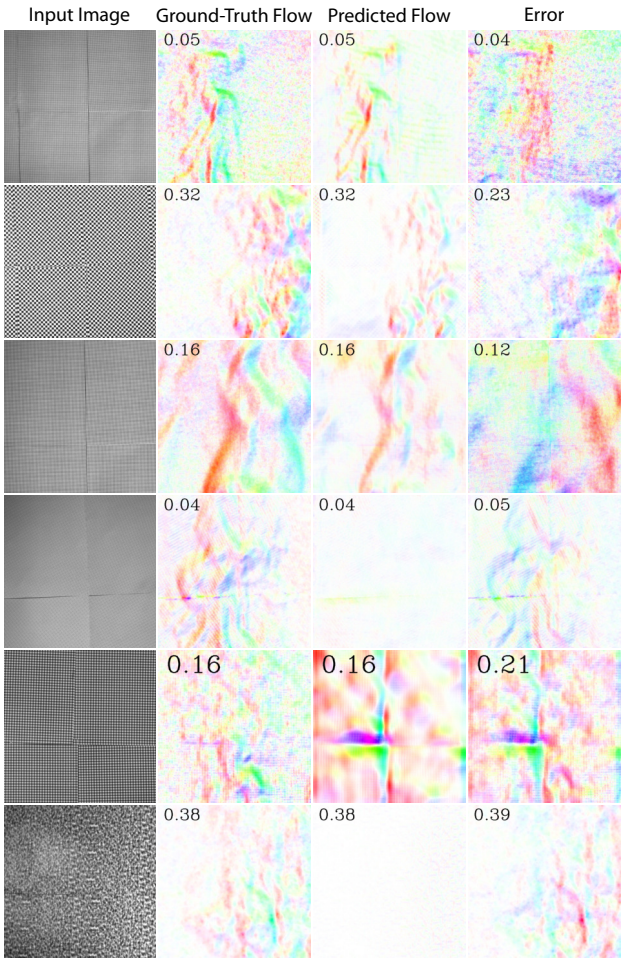

Fig. 7: Qualitative prediction results of the DYBS on previously unseen ROSCINE data (images containing real schlieren from a heat source). The maximum flow magnitude in pixels for each picture is indicated in the top left corner. For clarity, the same magnitude color scale is used for both the predicted and ground-truth flows.

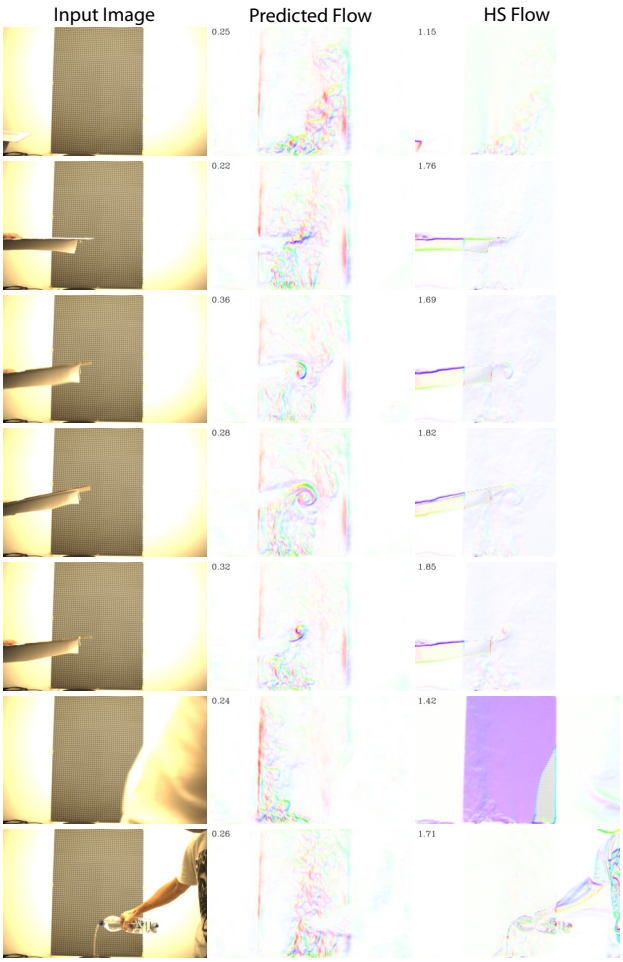

Fig. 8: Images with schlieren recorded in the indoor setup with different flow patterns than those observed during training. The flows are deflected by objects above the heat plate or sideways wind. For a sequence of frames refer to the accompanying video. We compare the DYBS predictions to the BOS optical flow.

the bottom two rows of Fig. 6, DYBS cannot detect the distortions but also does not hallucinate. However, even very little texture, such as a beach or a road in the same cases, is sufficient for DYBS.

The average magnitude prediction error on the ROSCINE data is significantly higher compared to the synthetically constructed ones at $0.034\,\mathrm{px}$. As evident based on the DYBS predictions in Fig. 7, there are two modes. The top three cases, where the background is similar to the ones observed during the training (same pattern but different distance between the camera and the background), are reasonably well predicted. However, DYBS struggles with the bottom three cases, where the background differs more, by either predicting zero or showing strong artifacts along the borders of the individual background sheets. Therefore, the success in detecting the schlieren for backgrounds from the Places dataset does not transfer to the various backgrounds used in the ROSCINE data collection.

Experiment 2: Flow Shape Variation. In our static indoor setup with the high-texture background, as shown in Fig. 2, we recorded a test subset of ROSCINE containing different flow patterns compared to those in the training set. We generated these different flow shapes using the same large heat plate used for the training subset, but with the flow disturbed by wind or objects deflecting the flow. We compare the CNN predictions to the BOS Horn-Schunck optical flow, where we used the average of the first 50 frames as the known background image, and show the results for select frames in Fig. 8. It is important to note that this data contains objects and flow patterns that are completely novel to the prediction network and do not appear in any of the training data. Nevertheless, DYBS manages to predict the new optical flow patterns well, such as the vortices at the plate tip, with minor artifacts at the border of the high-texture background and the objects. The flow quality from DYBS using a single image frame is arguably better than the flows from the two-frame BOS algorithm, which exhibits artifacts around the objects multiple times larger than the optical flow due to the schlieren. Nevertheless, we can still see that the BOS and DYBS flow patterns have similar shapes, demonstrating that the network generalizes

well to different flow patterns than those observed during training. Finally, the proposed method shows the highly desirable property of primarily returning the optical flow due to schlieren, rather than all elements of the scene motion that are extracted by the standard BOS optical flow.

Experiment 3: Moving Camera. We recorded images containing schlieren with a moving camera in an indoor setting with a high-texture background as well as outdoors with a natural background. The moving camera rules out using the traditional BOS algorithm to provide a reference flow, as it relies on a known undistorted background image, and introduces additional challenges such as motion blur and varying viewpoints. The DYBS predictions together with the input image for select frames are shown in Fig. 9. In the indoor setting the network was able to predict the optical flows from different viewpoints on the high-texture background. However, we also highlight some failure cases (rows 4 and 5). Row 5 shows an example from an outdoor setting where the heat source is several meters from the camera. Here the distances between the background, heat source and camera are very different to those used during training and therefore DYBS struggles to produce accurate predictions. These findings are consistent with our previous results and demonstrates a need for further investigation to improve generalization performance.

V. DISCUSSION AND LIMITATIONS

The ability to capture schlieren from a single image frame is a particularly valuable property, as it permits moving cameras in a way that was not previously possible. Remaining single-frame effects from camera motion, such as motion blur and changes in intensity due to aperture or exposure time variations, remain to be fully explored, but the results from this paper already show strong promise in this direction. Further, detecting schlieren with a moving camera from different viewpoints potentially allows 3D triangulation of the location of these flows to autonomously locate thermal updrafts or gas/heat leaks with a mobile robot.

Currently the results with the varying backgrounds on the Places Standard dataset do not transfer to unseen backgrounds captured with a camera. We reason that the low number of backgrounds (11) compared to the synthetic ones (1.8 million) in combination with different noise properties causes this performance difference. Generating a dataset of schlieren images through geometric-optics ray-tracing [35], [36] could help to generate a wider variety of flow pattern. Together with enhancing these images with photogrammetric noise, such as Gaussian noise, salt-and-pepper noise, motion blur, brightness, or contrast changes, a trained network might be able to better generalize to natural backgrounds captured with different cameras.

In our work we used single greyscale images as the input and extracted the ground-truth flows from each of them separately. Refraction indices differ slightly with varying wavelengths [37]. A standard optical camera captures the visible light in the range of $400\,\mathrm{nm}$ to $700\,\mathrm{nm}$ wavelength,

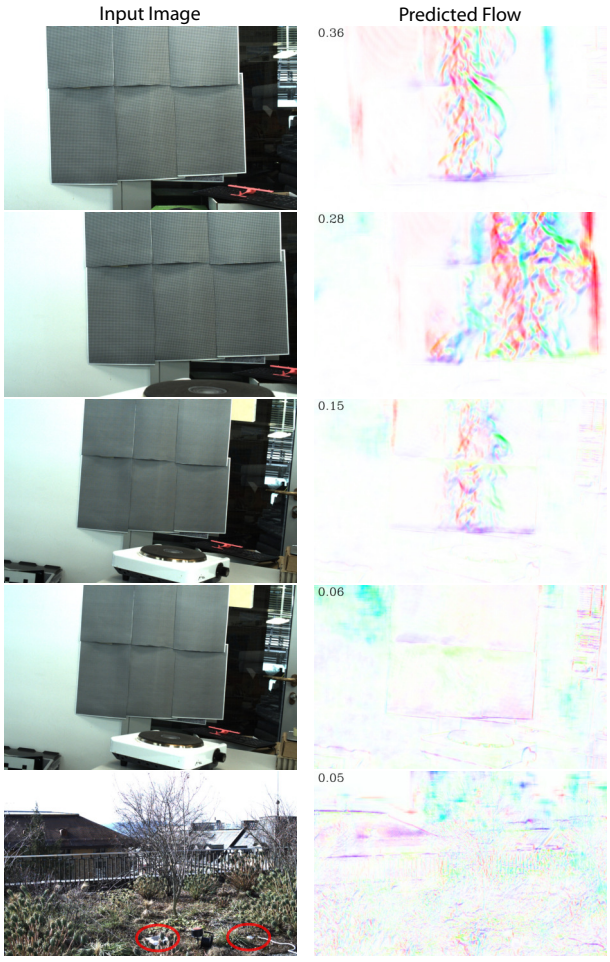

Fig. 9: Images with schlieren recorded with indoor and outdoor backgrounds and a moving camera and the corresponding DYBS predictions. In the outdoor scene the heat sources are highlighted by red circles.

thus for each channel the schlieren shapes should be unique. The neural network might be able to use this additional information to predict the optical flow more precisely if we extend the input to a multi-channel optical image.

VI. CONCLUSIONS

In this work we developed DYBS, a CNN-based approach, that is capable of predicting the sub-pixel image distortions due to schlieren from a single greyscale image. We recorded schlieren flow patterns in a controlled static indoor environment with a traditional BOS method and then trained the network with a mixture of real and synthetic samples by applying the extracted flows to new images. The resulting network learned to predict the sub-pixel flow patterns well with the random backgrounds from the Places Standard dataset and previously unseen flow patterns. While the current model does not generalize well to setups where the environments differ too much from the training images in terms of background texture and distances to the background and heat source, this should be addressable by collecting more diverse training data – the main avenue for future work.

REFERENCES

[1] G. Settles, "Colour-coding schlieren techniques for the optical study of heat and fluid flow," *International Journal of Heat and Fluid Flow*, vol. 6, no. 1, pp. 3–15, 1985.

[2] A. Okhotsimskii and M. Hozawa, "Schlieren visualization of natural convection in binary gas–liquid systems," *Chemical Engineering Science*, vol. 53, no. 14, pp. 2547–2573, 1998.

[3] Y. Cao, M. Dahari, I. Tlili, and A. Raise, "Investigation on the laminar flame speed of CH4/CO2/air mixture at atmospheric and high pressures using schlieren photography," *International Journal of Hydrogen Energy*, vol. 45, no. 55, pp. 31 151–31 161, 2020.

[4] P. J. Bruce and H. Babinsky, "Unsteady shock wave dynamics," *Journal of Fluid Mechanics*, vol. 603, pp. 463–473, 2008.

[5] Z. Akos, M. Nagy, S. Leven, and T. Vicsek, "Thermal soaring flight of birds and unmanned aerial vehicles," *Bioinspiration & Biomimetics*, vol. 5, no. 4, 2010.

[6] C. D. Cone, "Thermal soaring of birds," *American Scientist*, vol. 50, no. 1, pp. 180–209, 1962.

[7] C. M. Bishop, R. J. Spivey, L. A. Hawkes, N. Batbayar, B. Chua, P. B. Frappell, W. K. Milsom, T. Natsagdorj, S. H. Newman, G. R. Scott, J. Y. Takekawa, M. Wikelski, and P. J. Butler, "The roller coaster flight strategy of bar-headed geese conserves energy during himalayan migrations," *Science*, vol. 347, no. 6219, pp. 250–254, 2015.

[8] H. J. Williams, E. L. C. Shepard, M. D. Holton, P. A. E. Alarcón, R. P. Wilson, and S. A. Lambertucci, "Physical limits of flight performance in the heaviest soaring bird," *Proceedings of the National Academy of Sciences*, vol. 117, no. 30, pp. 17 884–17 890, 2020.

[9] M. Allen, "Autonomous soaring for improved endurance of a small uninhabitated air vehicle," in *43rd AIAA Aerospace Sciences Meeting and Exhibit*, 2005, p. 1025.

[10] P. Oettershagen, A. Melzer, T. Mantel, K. Rudin, T. Stastny, B. Wawrzacz, T. Hinzmann, S. Leutenegger, K. Alexis, and R. Siegwart, "Design of small hand-launched solar-powered UAVs: From concept study to a multi-day world endurance record flight," *Journal of Field Robotics*, vol. 34, no. 7, pp. 1352–1377, 2017.

[11] I. Guilliard, R. J. Rogahn, J. Piavis, and A. Kolobov, "Autonomous thermalling as a partially observable markov decision process," in *Robotics: Science and Systems*, 2018.

[12] G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski, and M. Vergassola, "Glider soaring via reinforcement learning in the field," *Nature Letters*, vol. 562, pp. 236–239, 2018.

[13] F. Achermann, A. Kolobov, D. Dey, T. Hinzmann, J. J. Chung, R. Siegwart, and N. Lawrance, "Multipoint: Cross-spectral registration of thermal and optical aerial imagery," in *Proceedings of the 2020 Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, J. Kober, F. Ramos, and C. Tomlin, Eds., vol. 155. PMLR, 16–18 Nov 2021, pp. 1746–1760.

[14] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[15] G. S. Settles, *Schlieren and shadowgraph techniques: visualizing phenomena in transparent media*. Springer Science & Business Media, 2001.

[16] G. Settles, *Toepler's Schlieren Technique*. Berlin, Heidelberg: Springer Berlin Heidelberg, 01 2001, pp. 39–75.

[17] M. Raffel, "Background-oriented schlieren (BOS) techniques," *Experiments in Fluids*, vol. 56, no. 3, pp. 1–17, 2015.

[18] N. T. Smith, J. T. Heineck, and E. T. Schairer, "Optical flow for flight and wind tunnel background oriented schlieren imaging," in *55th AIAA aerospace sciences meeting*, 2017, p. 0472.

[19] J. T. Heineck, D. W. Banks, N. T. Smith, E. T. Schairer, P. S. Bean, and T. Robillos, "Background-oriented schlieren imaging of supersonic aircraft in flight," *AIAA Journal*, vol. 59, no. 1, pp. 11–21, 2021.

[20] M. Raffel, J. T. Heineck, E. Schairer, F. Leopold, and K. Kindler, "Background-oriented schlieren imaging for full-scale and in-flight testing," *Journal of the American Helicopter Society*, vol. 59, no. 1, pp. 1–9, 2014.

[21] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[22] B. Atcheson, W. Heidrich, and I. Ihrke, "An evaluation of optical flow algorithms for background oriented schlieren imaging," *Experiments in fluids*, vol. 46, pp. 467–476, 2009.

[23] K. Perlin, "An image synthesizer," *ACM Siggraph Computer Graphics*, vol. 19, no. 3, pp. 287–296, 1985.

[24] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Scandinavian conference on Image analysis*. Springer, 2003, pp. 363–370.

[25] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[26] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, vol. 2. Morgan Kaufmann Publishers Inc., 1981, pp. 674–679.

[27] J. Wulff and M. J. Black, "Efficient sparse-to-dense optical flow estimation using a learned basis and layers," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[28] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "Deepflow: Large displacement optical flow with deep matching," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1385–1392.

[29] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[32] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8026–8037.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference for Learning Representations*, 2015.

[34] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual International Conference on Machine Learning*, 2009, pp. 41–48.

[35] G. Decker, R. Deutsch, W. Kies, and J. Rybach, "Computer-simulated schlieren optics," *Applied Optics*, vol. 24, no. 6, pp. 823–828, 1985.

[36] E. Luthman, N. Cymbalist, D. Lang, G. Candler, and P. Dimotakis, "Simulating schlieren and shadowgraph images from LES data," *Experiments in Fluids*, vol. 60, pp. 1–16, 2019.

[37] P. E. Ciddor, "Refractive index of air: new equations for the visible and near infrared," *Applied Optics*, vol. 35, no. 9, pp. 1566–1573, Mar 1996.