Enhancing Keypoint Detection in Thermal Images through Loss Function Optimisation and Model Evaluation

Gabriela Vdoviak, Tomyslav Sledevič

Vilnius Gediminas Technical University | Faculty of Electronics

INTRODUCTION

The application of thermal imaging for activity recognition has become crucial in several domains, including surveillance, healthcare, robotics, augmented reality, autonomous vehicles, behavioral analysis, and sports. These fields often operate under challenging conditions, such as low light environments, variations in illumination, and strict privacy concerns. Conventional RGB-based methods are frequently constrained by these limitations, whereas thermal imaging maintains performance in low-visibility environments and anonymizes identifiable features, making it suitable for privacy-sensitive tasks.

This research introduces a novel single-person thermal image dataset and aims to enhance keypoint detection performance in thermal images through the optimization of the loss function. In addition, this study evaluates different YOLOv8-Pose models in terms of keypoint detection accuracy and processing time.

MATERIALS AND METHODS

Dataset

A single person thermal image dataset of 1000 images (640×512 px) was collected using a Lynx L15 thermal monocular. The videos were captured in a real-world indoor environment at 25 fps and processed by extracting every fifth frame. Each frame was annotated with bounding boxes and 17 keypoints per person. The dataset captures multiple activities, including walking, standing, sitting, and other.

Fig 1. Samples of thermal images in dataset.

YOLOv8-Pose Model

For the experiments, we used the YOLOv8-Pose model. In the first experiment for loss function optimization, only the Nano model was trained with a batch size of 40 for 100 epochs to evaluate the impact on detection accuracy with different loss functions. In the second model evaluation experiment, we have used Nano, Small, Medium, Large, and Extra Large models to assess their performance in keypoint detection and processing capabilities. The batch size was adjusted according to the complexity of the model, ranging from 40 to 8 for 600 epochs. All experiments were performed on NVIDIA GeForce RTX 2060 GPU with 6 GB of RAM, and on the Ultralytics YOLOv8 framework using Python 3.9.17, PyTorch 2.1.2, and CUDA 11.8.

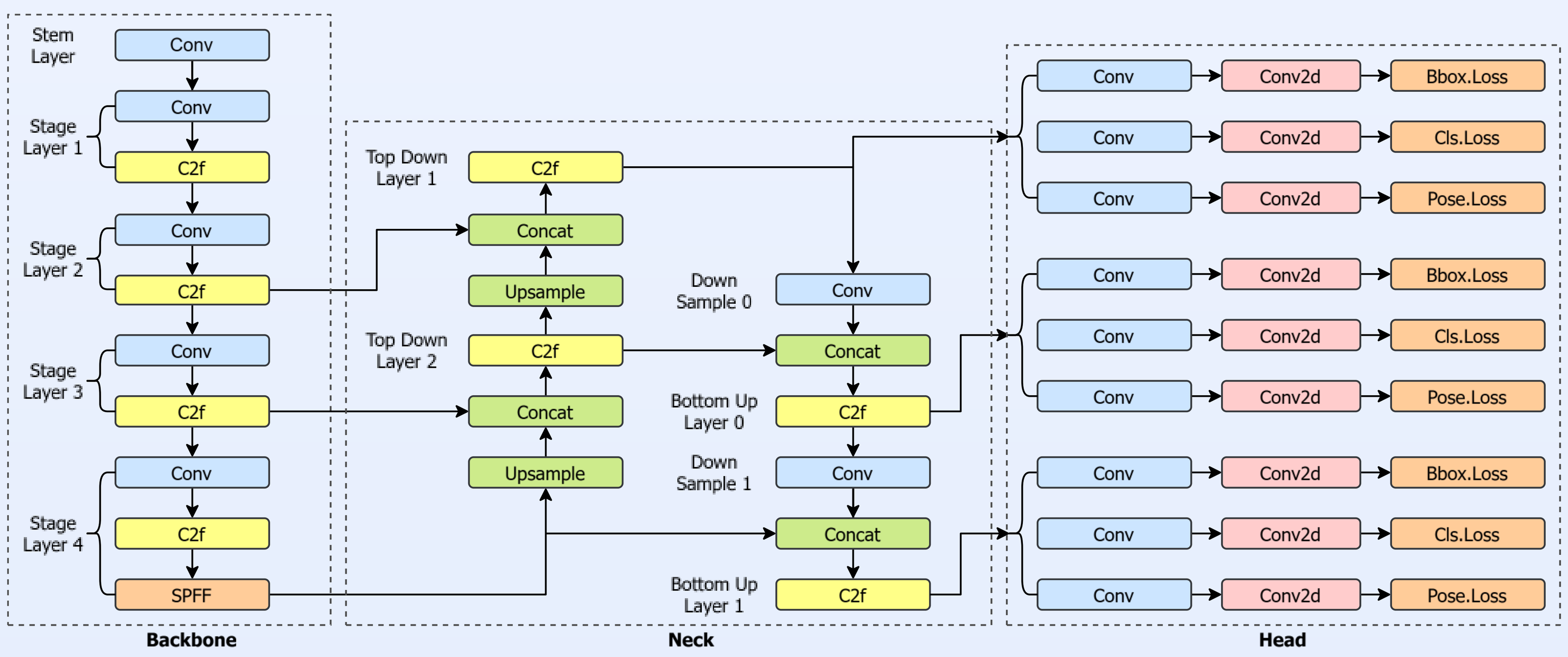

Fig 2. YOLOv8-Pose model architecture.

Loss Functions

To improve the accuracy of keypoint detection, several loss functions were used during training of the YOLOv8n-Pose model. The primary metric for evaluation was the Object Keypoint Similarity (OKS), which measures similarity between predicted and ground truth keypoints. In addition to the default loss function, the L_1 loss (Mean Absolute Error) and L_2 loss (Mean Squared Error) were examined to assess their impact on detection performance. These loss functions were calculated as follows:

$$KS_i = \exp\left(\frac{-d_i^2}{2s^2(2\sigma_i)^2}\right)$$

$$OKS = \frac{\sum_{i=1}^{17} KS_i \delta(v_i > 0)}{\sum_{i=1}^{17} \delta(v_i > 0)}$$

$$L_{OKS} = 1 - OKS$$

$$L_1 = \frac{\sum_{i=1}^{17} \left|y_i^{gt} - y_i^{pred}\right| \delta(v_i > 0)}{\sum_{i=1}^{17} \delta(v_i > 0)}$$

$$L_2 = \frac{\sum_{i=1}^{17} (y_i^{gt} - y_i^{pred})^2 \delta(v_i > 0)}{\sum_{i=1}^{17} \delta(v_i > 0)}$$

RESULTS

Table 1 summarizes the keypoint detection accuracy using mAP and OKS metrics. Training with the default L_{OKS} loss OKS achieved a keypoint similarity of 0.958 and a mAP of 0.966. The body only OKS improved by 1.9% due to the exclusion of the head keypoints. Replacement of L_{OKS} with L_1 loss increased mAP by 2.1% and OKS by 1.2%, while L_2 loss reduced both metrics by 1%. The optimal combination of L_{OKS} and L_1 increased keypoint similarity by 1.3% compared to L_{OKS} alone.

Table 1. Results of keypoint detection using different loss functions.

Loss Function	OKS	OKS (body)	mAP50-95
L_{OKS} (default)	0.958	0.957 (-0.1%)	0.966
L_1	0.970 (+1.2%)	0.977 (+1.9%)	0.987
$0.2L_{OKS} + 0.8L_1$	0.971 (+1.3%)	0.976 (+1.8%)	0.985
L_2	0.948 (-1%)	0.964 (+0.6%)	0.955
$0.2L_{OKS} + 0.8L_2$	0.954 (-0.4%)	0.965 (+0.7%)	0.968

Figure 3 illustrates predicted poses (blue) alongside ground truth (green). The orange dots represent the ground truth keypoints, while the red dots indicate predictions. The highest OKS scores are achieved when the person's body is fully visible and facing the camera. In contrast, scores decrease when the body is rotated or part of the keypoints are occluded. Therefore, depending on the application specifics and the requirements for pose detection accuracy, the dataset needs to be augmented by images with occluded or missing keypoints.

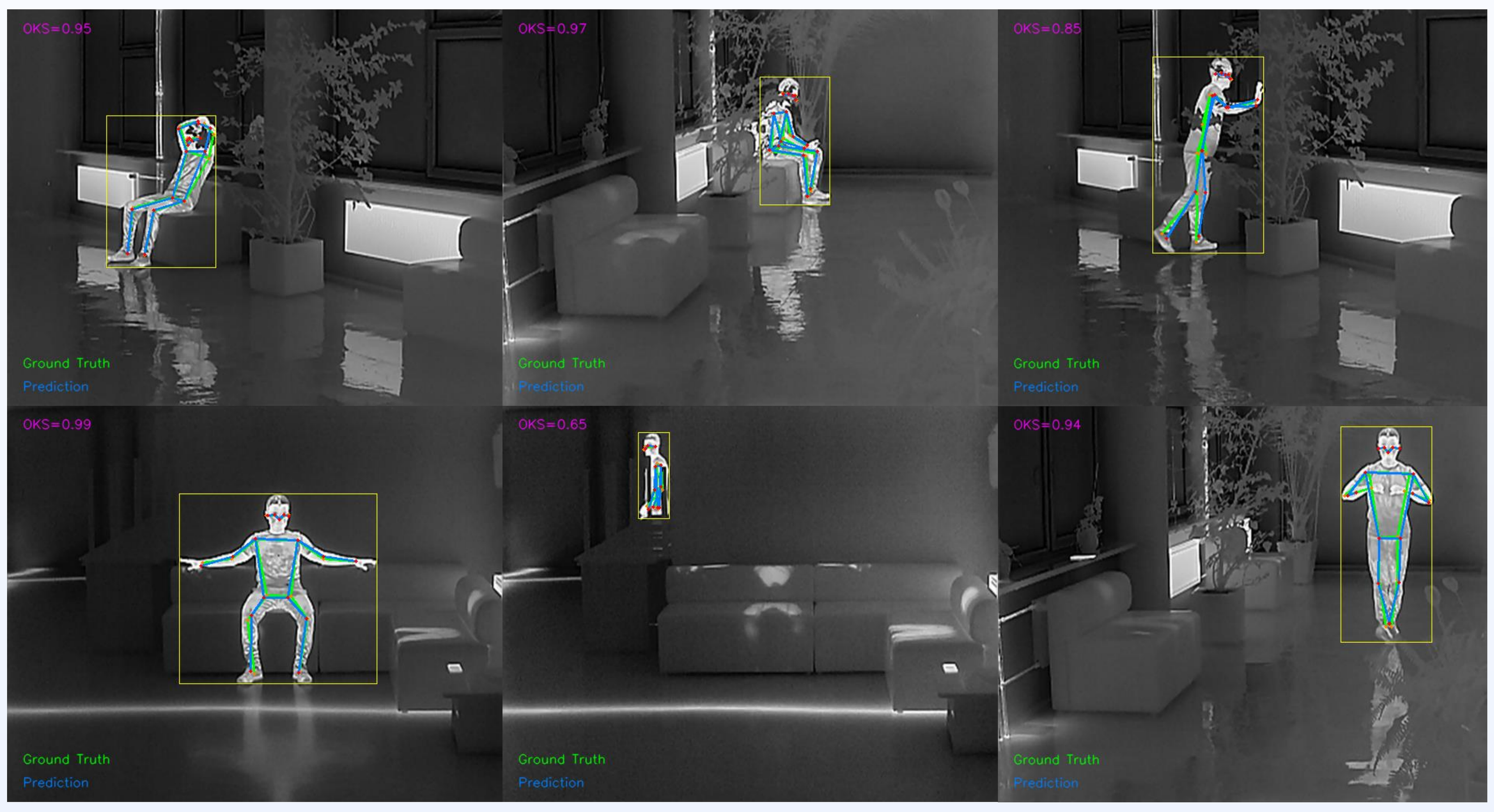

Fig 3. Keypoint detection results in thermal images with OKS metric, ground truth and predicted keypoints.

Table 2 summarizes the keypoint detection results for different YOLOv8-Pose models using the OKS evaluation metric. The difference in OKS between the Nano and Extra-Large models is only 0.9%, despite the latter having 21 times more parameters. This indicates that increasing the size of the model does not result in substantial improvements in precision. Furthermore, the Nano model processes frames 3.8 times faster than the Extra-Large model, making Nano and Small models more suitable for real-time applications.

Table 2. Results of keypoint detection after training of YOLOv8-Pose models.

	Nano	Small	Medium	Large	Extra Large
Parameters, M	3.3	11.6	26.4	44.4	69.4
Batch size	40	24	12	10	8
OKS, %	95.6	95.8	96.2	96.3	96.5
Time per frame, ms	10.4	13.8	24.7	26.6	39.9

CONCLUSIONS

This research demonstrates the potential of thermal imaging for keypoint detection by optimizing the loss function and evaluating different YOLOv8-Pose models in terms of accuracy and processing capabilities. The findings indicate that the combination of L_{OKS} and L_1 loss functions achieved the highest keypoint detection accuracy of 0.97, improving OKS by 1.3% compared to the default loss function. Among the YOLOv8-Pose models, the Nano and Small models demonstrated the optimal balance between accuracy and computational efficiency. These findings highlight the importance of both loss function optimisation and model selection in enhancing keypoint detection for thermal images, with implications for real-time applications in diverse domains.

Further research could investigate strategies for data augmentation to enhance model robustness in challenging scenarios and the integration of other sensors with thermal imaging.

15th Conference Data Analysis Methods for Software Systems
November 28 – 30, 2024, Druskininkai, Lithuania