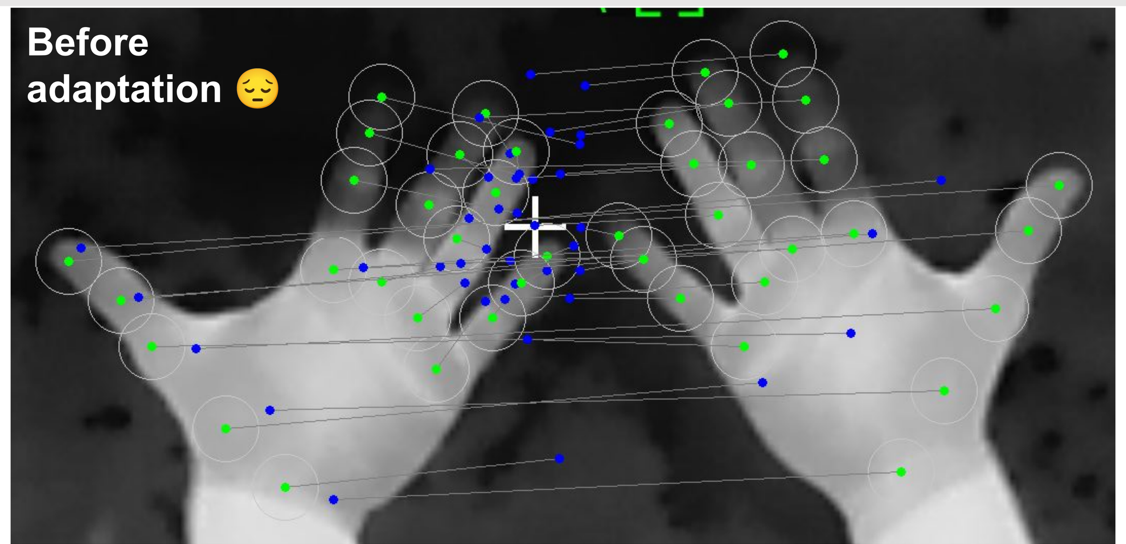


Domain Adaptation for Enhancing Visual Hand Landmark Prediction AI in Infrared Imaging

Unsupervised domain adaptation with AdaBN, Deep CORAL, and SSA Improves Keypoint Detection by 11% on InterWild model trained on RGB, when tested on IR dataset

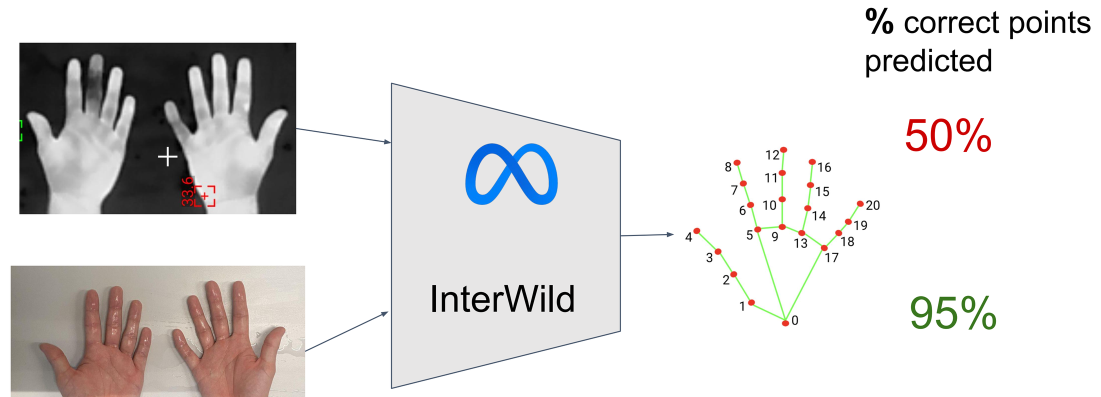
Vladimir Sachkov (v.sachkov@student.tudelft.nl)

Responsible professor: Jan van Gemert
Supervisors: Zhi-Yi Lin , Thomas Markhorst

Code available at: https://github.com/EraChanZ/RP

Problem & Motivation

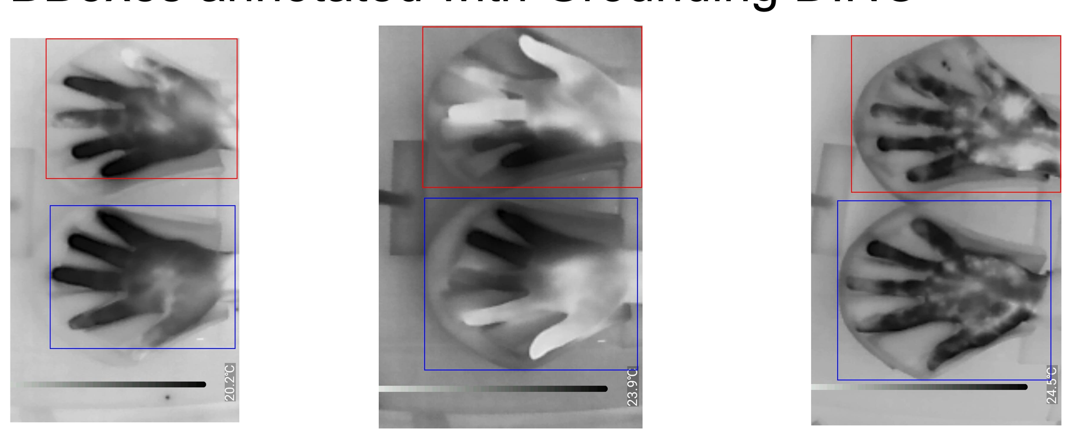
Problem: RGB-trained models fail on IR due to domain shift (thermal vs. visual features).
Medical Need: Early leprosy detection requires precise temperature measurement at hand joints.
Challenge: Limited labeled IR data (only 80 labeled images and ≈ 4500 unlabeled).



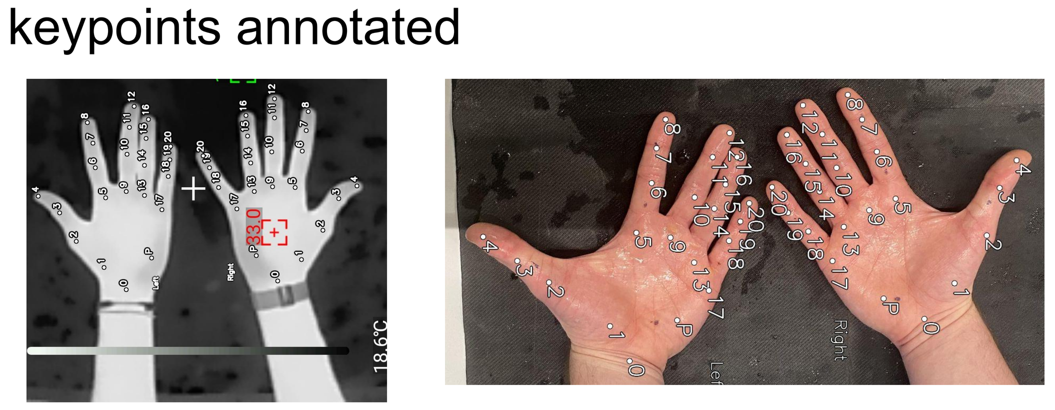
InterWild performs 45% less accurate on IR compared to RGB domain

Data

1 FPS sampled from IR videos ≈ **5000** frames
BBoxes annotated with Grounding DINO



80/80 RGB/IR images, manually collected + keypoints annotated



Source domain (RGB datasets)

InterHand26M
≈ **2.6 * 10^6** images

COCO-Wholebody
≈ **1.3 * 10^5** images

Methodology

AdaBN

- **Main idea:** Batch normalization statistics (running mean and variance) carry information about the domain distribution.
- **Implementation:**
  - Set batch size
  - Set momentum for BN layers
  - Perform inference on all target domain, while keeping BN layers in train mode.

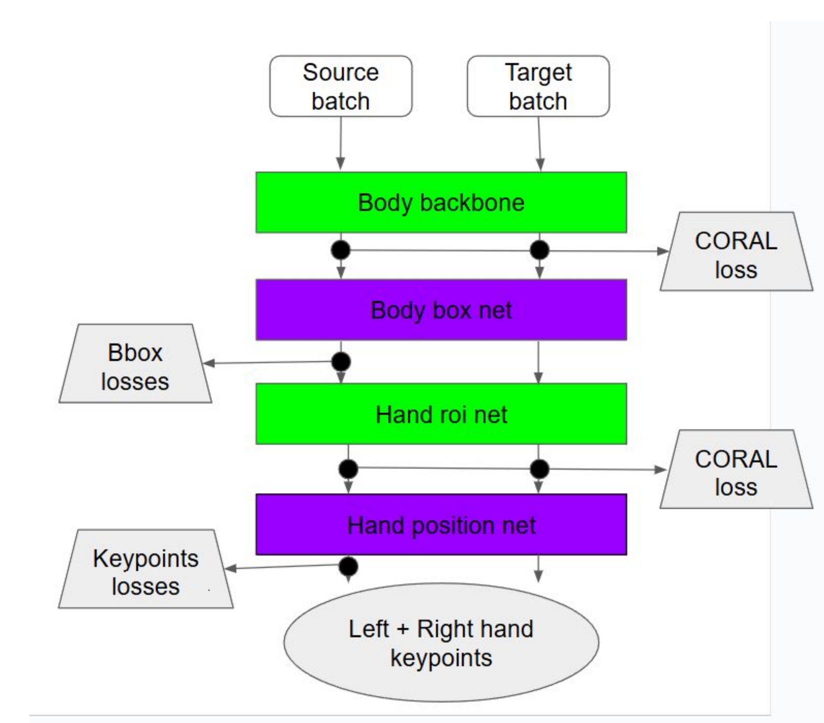
DeepCORAL

- **Main idea:** Introduce CORAL loss to a new regularized loss criteria for training on both source (RGB) and target (IR) datasets to maximize feature alignment

$$\mathcal{L}_{\text{body}} = \mathcal{L}_{\text{bbox}} + \lambda \ell_{\text{CORAL}}^{\text{body}}$$

λ - CORAL weight

$$\mathcal{L}_{\text{hand}} = \mathcal{L}_{\text{kp}} + \lambda \ell_{\text{CORAL}}^{\text{hand}}$$



SSA (Test-time Adaptation for Regression by Subspace Alignment)

- **Main idea:** Features of regression models have low subspace dimensionality -> naive feature alignment is unstable due to many dimensions having zero variance
- **Implementation:**
  - Calculate means and covariance matrices on source domain (COCO + InterHand)
  - Fine-tune InterWild on target domain only using alignment loss, in advance projecting batches on subspace formed by top-K eigenvectors of source domain
  -

Evaluation framework

A Python library was designed to assess model performance using:
- **IoU** (intersection over union) Measures bounding box localization accuracy via overlap ratio.
- **PCK** (Percentage of correct keypoints)
  - **PCK@0.05** Evaluates keypoint correctness with a threshold of 5% of image size.
  - **APCK** Adaptive threshold based on ground truth keypoint spacing for anatomical precision.

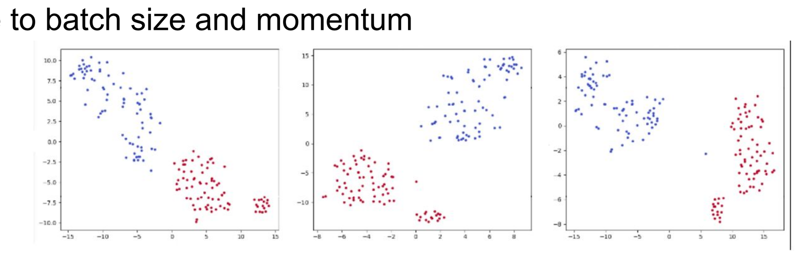
The framework combines quantitative metrics with visualizations (box/keypoint overlays) to validate model robustness across scales and anatomical variations.

Experiments

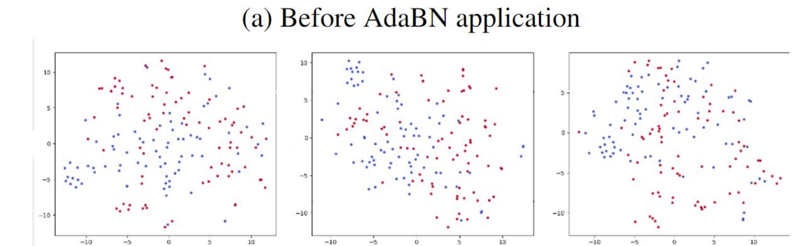
AdaBN

- Hyperparameter search performed over batch size, BN layer momentum, inclusion of frames sampled from videos

Observations:
- Inclusion of data, apart from data model is tested on only decreased performance
- Moderate **+2-4%** PCK improvement
- Highly sensitive to batch size and momentum



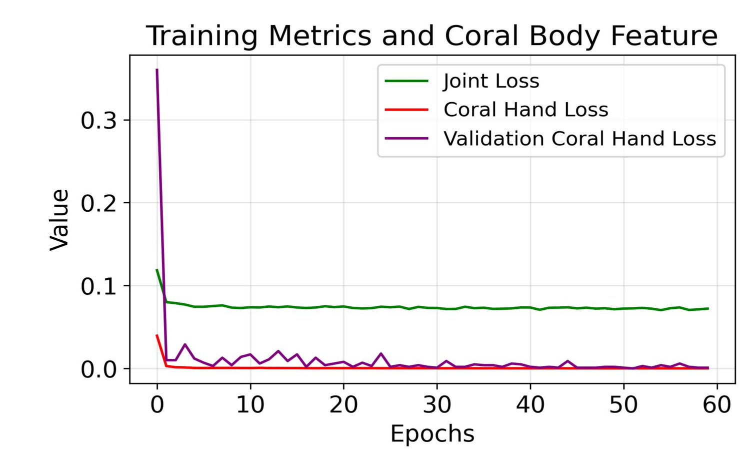
(a) Before AdaBN application



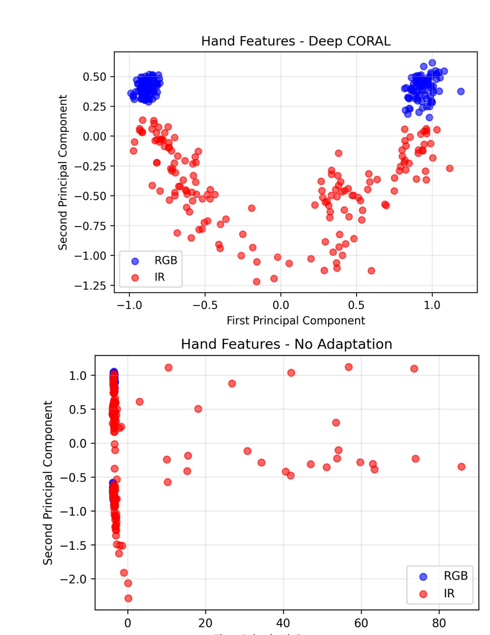
(b) After AdaBN application

DeepCORAL

- Focused on Hand_roi_net training using infrared dataset with annotated hand bounding boxes. Source domain was randomly sampled to align with IR dataset for every epoch. **4-6%** PCK improvement



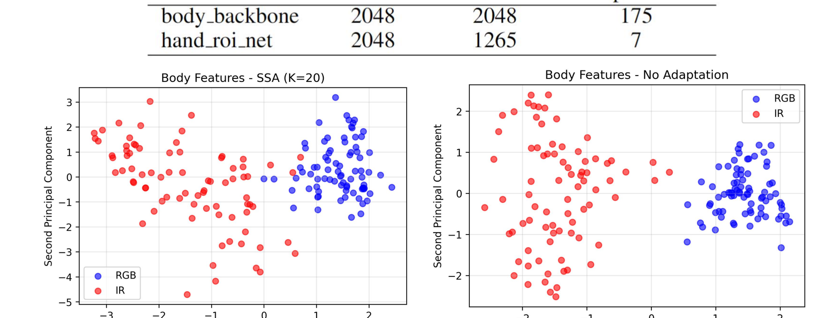
Observations
- No matter which learning rate, or coral_weight was chosen, coral_loss would converge very quickly after first few epochs. Two main reasons:
  - IR dataset uncomparably smaller, and much less diverse compared to source dataset.
  - Average Adaptive pooling was utilized to reduce hand feature dimensionality from [2048, 8, 8] -> [2048, 1, 1], hence we can also not see perfect feature alignment on visualization graphs



SSA

- Modular training of the model (separately hand_roi_net and body_backbone) allowed for a hyperparameter tuning on a batch size of 64. Only training of body_backbone improved the baseline by + **3-5%** PCK, despite body_backbone having full valid dimensions compared to hand_roi_net. Feature visualization shows only minimal alignment.

Module	Total dims	Valid dims	Subspace dims
body_backbone	2048	2048	175
hand_roi_net	2048	1265	7

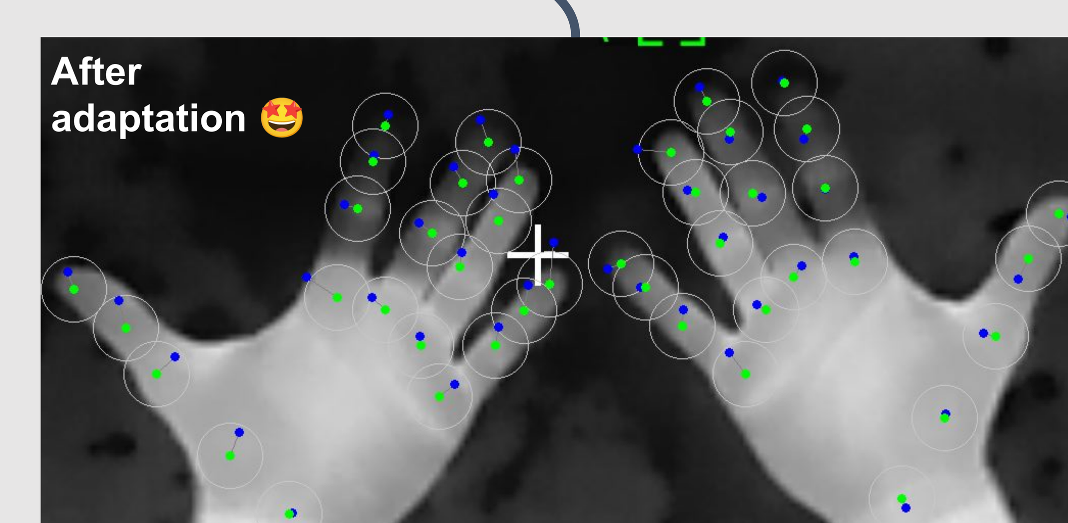


Ensembling & Conclusion

AdaBN is broadly compatible with complementary methods, requiring only test-data batch statistic recalibration pre-inference. While inducing strong feature alignment, integration yielded modest gains (+0.5–1.5% PCK), indicating BN updates alone insufficiently resolve pose estimation. InterWild's modularity enabled direct integration of optimal **SSA** (body_backbone) and **DeepCORAL** (hand_roi_net) checkpoints, achieving an **11%** PCK improvement—demonstrating pretrained component integration outperforms standalone statistical adaptation.

Method	Full Dataset			Cleaned Dataset		
	IOU	PCK@0.05	APCK	IOU	PCK@0.05	APCK
InterWild (baseline)	0.625	0.656	0.498	0.718	0.777	0.650
AdaBN	0.654	0.676	0.524	NA	NA	NA
Deep CORAL (hand)	0.675	0.741	0.585	0.763	0.840	0.705
SSA (body)	0.646	0.693	0.543	0.710	0.753	0.635
Deep CORAL + AdaBN (best)	0.682	0.756	0.596	NA	NA	NA
SSA (body) + AdaBN (best)	0.656	0.704	0.553	NA	NA	NA
SSA (body) + DeepCORAL (hand)	0.717	0.780	0.617	0.787	0.874	0.727

Despite spatial reductions, small scale unlabeled IR dataset, and modular based training, deep learning methods such as **SSA** and **DeepCORAL** have shown to be effective in improving accuracy of keypoint detection, as well as improving feature alignment of internal layers in InterWild architecture for RGB and IR domains.



Future work

- Employ the complete InterWild architecture for training and systematic hyperparameter optimization applied to both SSA and DeepCORAL methods, leveraging enhanced computational resources to facilitate the process.
- construct large-scale synthetic hands dataset using infrared camera simulation in combination with 3d physics engines
- Implement "Regressive Domain Adaptation for Unsupervised Keypoint Detection" Zhang et al. (2021) arXiv:2103.06175