Deep Weakly Supervised Domain Adaptation for Pain Localization in Videos

Gnana Praveen R Eric Granger Patrick Cardinal

November 19 2020

Laboratoire d'imagerie, de vision et d'intelligence artificielle (LIVIA),
École de technologie supérieure, Montréal, Canada

Presentation Overview

1. Introduction

Presentation Overview

1. Introduction
2. Background on Weakly-Supervised Methods for Domain Adaptation

Presentation Overview

Presentation Overview

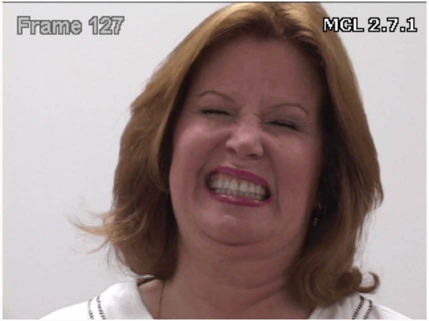
1. Introduction
2. Background on Weakly-Supervised Methods for Domain Adaptation
3. Proposed Approach
4. Results and Discussion

Presentation Overview

1. Introduction

Pain Detection in Health Care

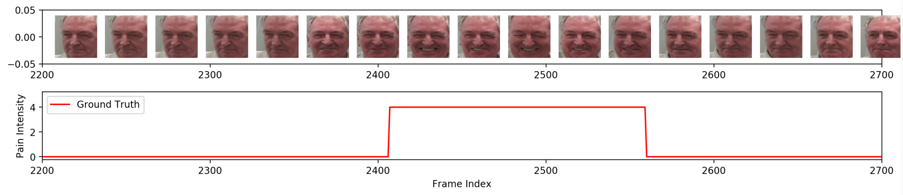
- Pain is a primary symptom of any malfunctioning in our body.
- Need for uninterrupted and contact-free monitoring of patients.
- Facial expressions convey significant information for pain detection.



Images taken from UNBC-McMaster pain dataset [Lucey et al., 2011]

Challenges in Pain Intensity Estimation

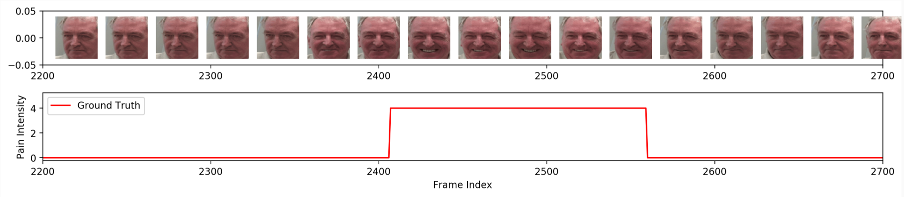
- Labeling of pain intensity levels in video frames is very expensive.



Images taken from UNBC-McMaster pain dataset [Lucey et al., 2011]

Challenges in Pain Intensity Estimation

- Labeling of pain intensity levels in video frames is very expensive.



- Level ambiguity due to subtle differences across intensities (pain and non-pain frames at left and right respectively)

Images taken from UNBC-McMaster pain dataset [Lucey et al., 2011]

Challenges in Pain Intensity Estimation

- High inter-subject variability (subject identity bias)

Images taken from Recola [Ringeval et al., 2013] and UNBC-McMaster [Lucey et al., 2011]

Challenges in Pain Intensity Estimation

- High inter-subject variability (subject identity bias)

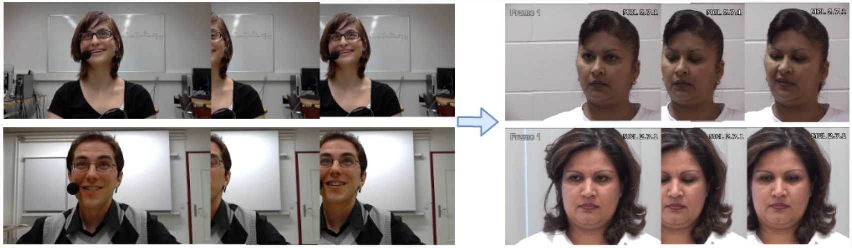
- Domain shift affected by several factors such as pose, occlusion and out-of plane head movements

Images taken from Recola [Ringeval et al., 2013] and UNBC-McMaster [Lucey et al., 2011]

2. Background on Weakly-Supervised Methods for Domain Adaptation

Why Domain Adaptation?

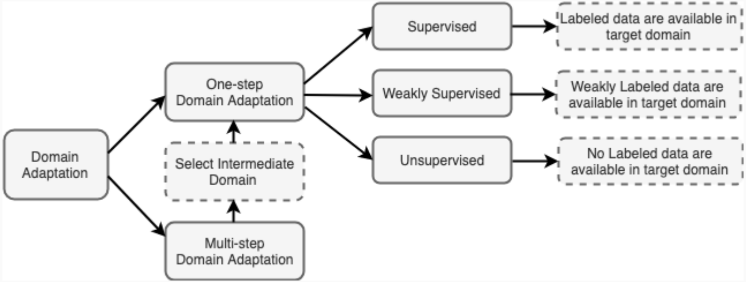
- Deep learning models are data hungry methods that requires millions of labeled examples.
- Models do not generalize well to new domains i.e., fail to cope with domain differences.

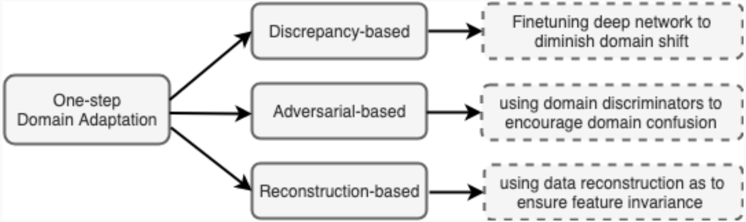


Source Dataset (Recola)   Target Dataset (UNBC-McMaster)

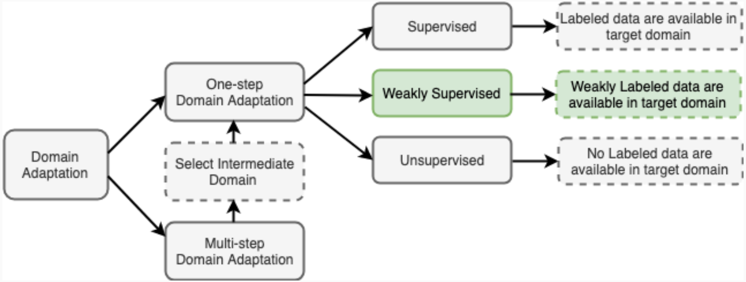
Images taken from Recola [Ringeval et al., 2013] and UNBC-McMaster [Lucey et al., 2011]

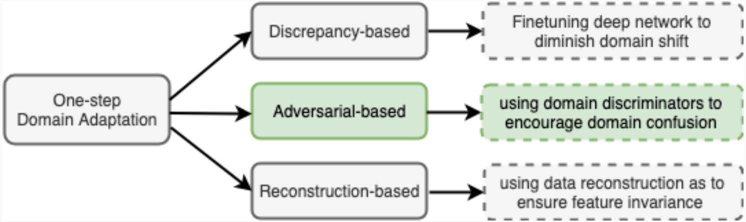
Domain Adaptation Approaches



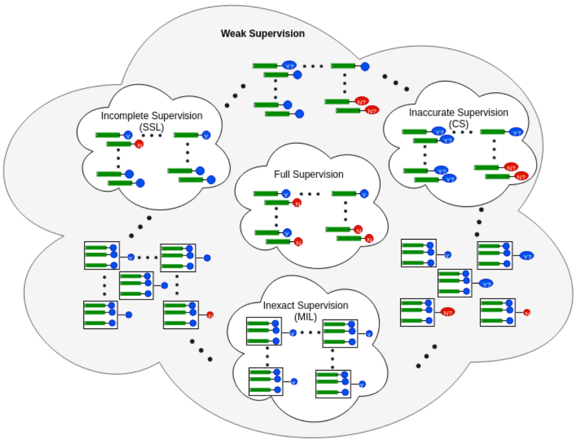


Domain Adaptation Approaches



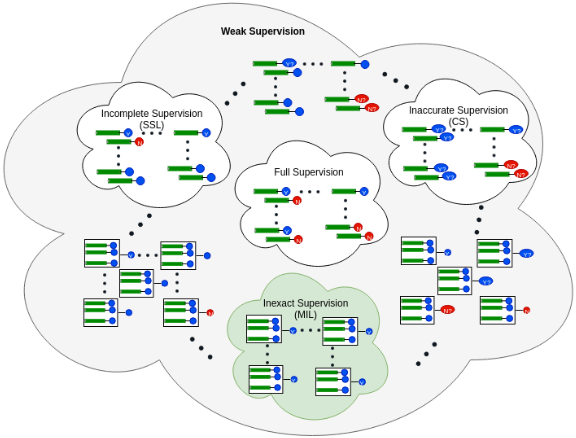


Weakly Supervised Learning (WSL)



Bars denote feature vectors; red/blue marks labels; “?” implies label may be inaccurate. Intermediate subgraphs depict some situations with mixed types of weak supervision. [Zhou, 2018]

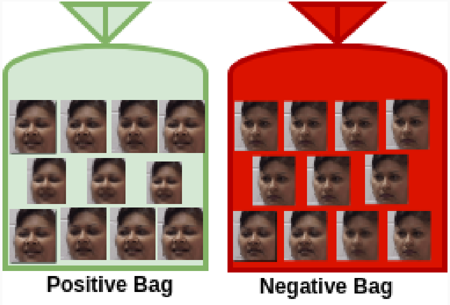
Weakly Supervised Learning (WSL)



Bars denote feature vectors; red/blue marks labels; “?” implies label may be inaccurate. Intermediate subgraphs depict some situations with mixed types of weak supervision. [Zhou, 2018]

Multiple Instance Learning (MIL) for Facial Expressions

- Bag denotes video sequence with multiple frames (instances).
- Positive Bag corresponds to video sequence with atleast one expression frame.
- Negative Bag corresponds to video sequence with only neutral frames.



MIL Approaches for Pain Estimation

- [Sikka et al., 2014] extracted spatiotemporal features (SIFT) and proposed pain localization and classification using MIL assuming video sequence as bag and subsequence as instance.

  $\longrightarrow$ Does not provide pain intensity levels for individual frames.

MIL Approaches for Pain Estimation

- [Sikka et al., 2014] extracted spatiotemporal features (SIFT) and proposed pain localization and classification using MIL assuming video sequence as bag and subsequence as instance.

  $\longrightarrow$ Does not provide pain intensity levels for individual frames.

- [Ruiz et al., 2018] proposed multi-instance dynamic ordinal random fields by modeling the relation between ordinal levels and given observation (frames) using normal distribution.

  $\longrightarrow$ Does not efficiently capture the spatiotemporal relationships.

Limitations of SOA Approaches

- The challenges of domain shift is not explored in the context of MIL.

Limitations of SOA Approaches

- The challenges of domain shift is not explored in the context of MIL.
- Regression of pain expressions is an under-researched problem.

Limitations of SOA Approaches

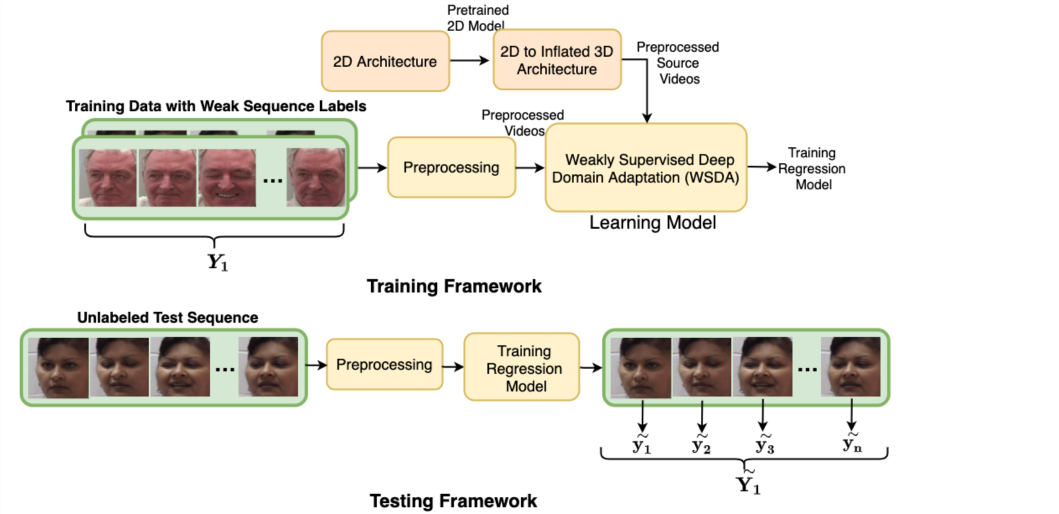
- The challenges of domain shift is not explored in the context of MIL.
- Regression of pain expressions is an under-researched problem.
- The potential of deep learning models is not explored with MIL due to lack of representative data.

Limitations of SOA Approaches

- The challenges of domain shift is not explored in the context of MIL.
- Regression of pain expressions is an under-researched problem.
- The potential of deep learning models is not explored with MIL due to lack of representative data.
- The temporal dynamics of pain expressions is not effectively captured in the existing approaches.

3. Proposed Approach

Overall Framework



Weakly Supervised Deep Domain Adaptation (WSDA)

- The training mechanism has four major modules : feature extractor, source loss, target loss and domain loss.
- 3D CNN is used for capturing spatiotemporal features and adversarial domain adaptation is deployed to optimize the domain loss.
- Source and target losses are computed using full source labels and weak target labels respectively.

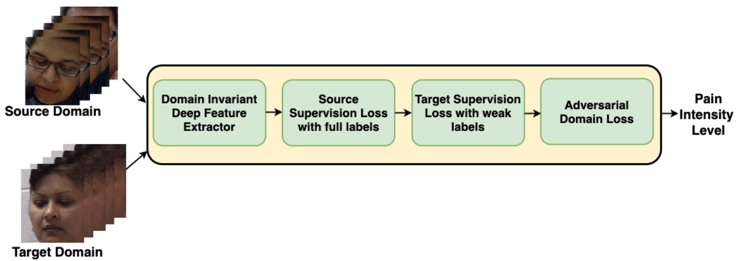
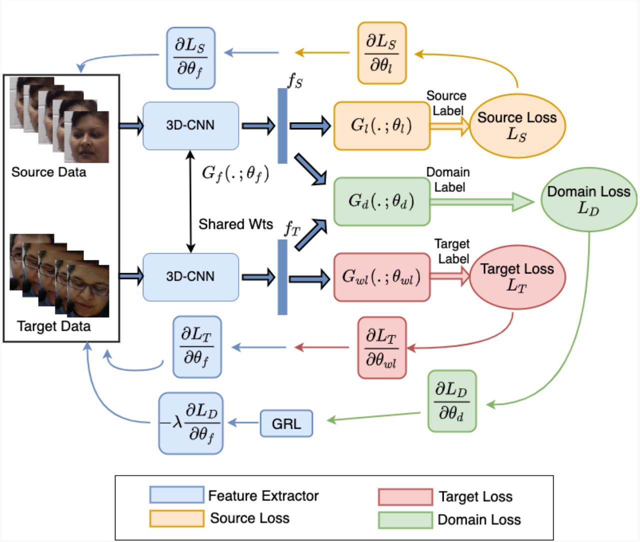


Figure: Block Diagram of the proposed approach

Weakly Supervised Deep Domain Adaptation (WSDA)



Weakly Supervised Deep Domain Adaptation (WSDA)

Source Loss

$$L_S = \frac{1}{N_s} \sum_{\substack{i=1 \\ d_i=0}}^{N_s} \sum_{j=1}^{n_i} ((G_l(G_f(x_i^j)) - y_i^j))^2$$

x_i^j : j$^{\text{th}}$ frame of i$^{\text{th}}$ sequence in source domain
y_i^j : label of j$^{\text{th}}$ frame of i$^{\text{th}}$ sequence
N_S: # of sequences in source domain

Weakly Supervised Deep Domain Adaptation (WSDA)

Source Loss

$$L_S = \frac{1}{N_s} \sum_{\substack{i=1 \\ d_i=0}}^{N_s} \sum_{j=1}^{n_i} ((G_l(G_f(x_i^j)) - y_i^j))^2$$

x_i^j : $\mathrm{j^{th}}$ frame of $\mathrm{i^{th}}$ sequence in source domain

y_i^j : label of $\mathrm{j^{th}}$ frame of $\mathrm{i^{th}}$ sequence

N_S: # of sequences in source domain

Target Loss

$$L_T = \frac{1}{N_T} \sum_{\substack{i=1 \\ d_i=1}}^{N_T} \sum_{j=1}^{n_i} ((G_l(G_f(X_i)) - Y_i))^2$$

X_i : $\mathrm{i^{th}}$ sequence of target domain

Y_i : Weak label of $\mathrm{i^{th}}$ sequence

N_T: # of sequences in target domain

Weakly Supervised Deep Domain Adaptation (WSDA)

Domain Loss

$$L_d = \frac{1}{N_s+N_T} \sum_{\substack{i=1 \\ d_i=0,1}}^{N_s+N_T} \sum_{j=1}^{n_i} [-d_i^j \log \left(G_d(G_f(x_i^j)) \right)$$

$$- (1\text{-}d_i^j) \log \left(1 - G_d(G_f(x_i^j))\right)]$$

x_i : j^{th} of i^{th} sequence of source or target domain

d_i: domain label of i^{th} sequence

Weakly Supervised Deep Domain Adaptation (WSDA)

Domain Loss

$$L_d = \frac{1}{N_s+N_T} \sum_{\substack{i=1 \\ d_i=0,1}}^{N_s+N_T} \sum_{j=1}^{n_i} [-d_i^j \log \left(G_d(G_f(x_i^j)) \right)$$

$$- (1\text{-}d_i^j) \log \left(1 - G_d(G_f(x_i^j))\right)]$$

x_i : jth of ith sequence of source or target domain

d_i: domain label of ith sequence

Overall Loss

$$L = L_S + L_T - \lambda L_d$$

λ : trade-off parameter between domain loss and prediction loss

4. Results and Discussion

Results with Baseline Training Models

Training Scenario	PCC ↑	MAE ↓
Supervised (source data only)	0.295	1.630
Supervised (target data only)	0.447	0.804
Supervised (source $\cup$ target)	0.612	0.543
Unsupervised DA	0.413	0.874
WSDA (ours)	0.676	0.774
Supervised DA	0.812	0.454

Table: PCC and MAE performance of I3D model trained under different scenarios.

Results with Varying Sequence Lengths

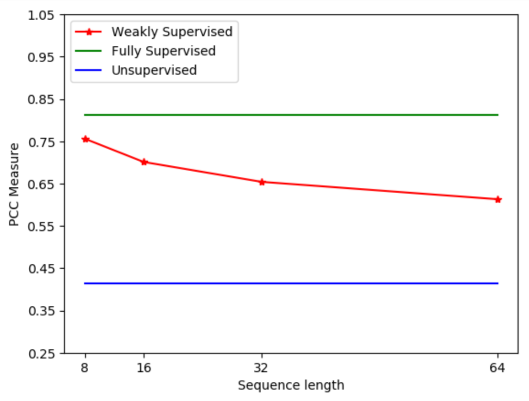


Figure: PCC accuracy of I3D model trained with deep WSDA levels with decreasing level of weak supervision on target videos.

Visualizations of Pain Localization on two subjects

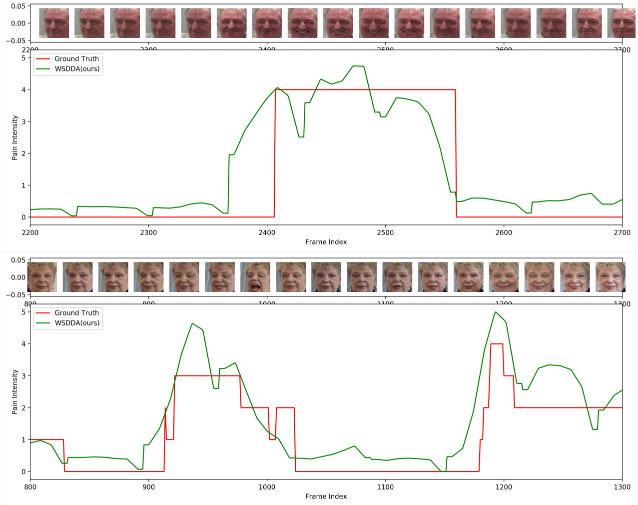


Figure: Visualization of pain localization on two different subjects. From top to bottom: scenario where ground truth (GT) shows no pain, but our deep WSDA approach correctly localizes pain. Scenario with multiple peaks of expressions

Comparison with state-of-the-art approaches

Method	Frame-level			Sequence-level		
	PCC ↑	MAE ↓	ICC ↑	PCC ↑	MAE ↓	ICC ↑
Weakly-Supervised						
MIR [Hsu et al., 2014]	0.350	0.840	0.240	0.63	0.940	0.630
MILBOOST [Sikka et al., 2014]	0.280	1.770	0.110	0.380	1.700	0.380
MI-DORF [Ruiz et al., 2018]	0.400	0.190	0.460	0.670	0.800	0.660
Deep WSDA (ours)	**0.630**	**0.714**	**0.567**	**0.828**	**0.647**	**0.762**
Semi-Supervised						
BORMIR [Zhang et al., 2018]	0.605	0.821	0.531	-	-	-
Fully Supervised						
LSTM [Rodriguez et al., 2018]	0.780	0.500	-	-	-	-
SCN [Tavakolian and Hadid, 2019]	0.920	0.320 (MSE)	0.750	-	-	-

Table: PCC, MAE and ICC performance of proposed and state-of-art methods.

5. Conclusion

Conclusion

- We propose a deep WSDA model that can significantly improve the performance of the system.

Conclusion

- We propose a deep WSDA model that can significantly improve the performance of the system.
- Domain Adaptation is explored for pain intensity estimation with weak labels.

Conclusion

- We propose a deep WSDA model that can significantly improve the performance of the system.
- Domain Adaptation is explored for pain intensity estimation with weak labels.
- Deep models are exploited for efficient feature extraction of the pain expressions.

Conclusion

- We propose a deep WSDA model that can significantly improve the performance of the system.
- Domain Adaptation is explored for pain intensity estimation with weak labels.
- Deep models are exploited for efficient feature extraction of the pain expressions.
- An extensive set of experiments are conducted with various baselines using various combinations of source and target domains.

References

Hsu, K., Lin, Y., and Chuang, Y. (2014).
Augmented multiple instance regression for inferring object contours in bounding boxes.
IEEE Trans. on Image Processing, 23(4):1722–1736.

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., and Matthews, I. (2011).
Painful data: The unbc-mcmaster shoulder pain expression archive database.
In *FG*.

Ringeval, F., Sonderegger, A., Sauer, J., and Lalanne, D. (2013).
Introducing the recola multimodal corpus of remote collaborative and affective interactions.
In *FG*.

Rodriguez, P., Cucurull, G., Gonzàlez, J., Gonfaus, J. M., Nasrollahi, K., Moeslund, T. B., and Roca, F. X. (2018).
Deep pain: Exploiting long short-term memory networks for facial expression classification.
IEEE Trans. on Cybernetics, pages 1–11.

Ruiz, A., Rudovic, O., Binefa, X., and Pantic, M. (2018).
Multi-instance dynamic ordinal random fields for weakly-supervised facial behavior analysis.
CoRR, abs/1803.00907.

Ruiz, A., Rudovic, O., Binefa, X., and Pantic, M. (2018).
Multi-instance dynamic ordinal random fields for weakly supervised facial behavior analysis.
IEEE Trans. on Image Processing, 27(8):3969–3982.

Sikka, K., Dhall, A., and Bartlett, M. S. (2014).
Classification and weakly supervised pain localization using multiple segment representation.
Image and Vision Computing, 32(10):659 – 670.

Tavakolian, M. and Hadid, A. (2019).
A spatiotemporal convolutional neural network for automatic pain intensity estimation from facial dynamics.
Int. Journal of Computer Vision, 127:1413 – 1425.

Zhang, Y., Zhao, R., Dong, W., Hu, B., and Ji, Q. (2018).
Bilateral ordinal relevance multi-instance regression for facial action unit intensity estimation.
In *CVPR*.

Zhou, Z.-H. (2018).
A brief introduction to weakly supervised learning.
National Science Review, 5(1):44–53.