# Towards Explaining Expressive Qualities in Piano Recordings: Transfer of Explanatory Features via Acoustic Domain Adaptation

**Shreyan Chowdhury, Gerhard Widmer**

Institute of Computational Perception
Johannes Kepler University Linz

JOHANNES KEPLER UNIVERSITY LINZ

Institute of Computational Perception

## INTRODUCTION

Acoustic features learnt from smaller datasets may not generalise well to specific domains. In this work, we explore this problem for *mid-level* features and provide a strategy to transfer these features to the domain of solo piano recordings.

### Why Transfer Mid-level Features?

These features are based on perceptual and intuitive musical qualities that have been shown to be effective in explaining musical emotion [1].

### Approach:

Two-step approach:

1. Gradient-based unsupervised domain adaptation (UDA)
2. Teacher-student mediated refinement

### Model Architecture

Receptive-Field regularised ResNet (RF-ResNet): Modified ResNet with smaller CNN kernels and fewer layers. This architecture has been shown to improve generalisation in audio and musical applications.

### Modeling Expressive Performance Descriptors

We use our models to extract average mid-level features for piano performance recordings and fit them to four dimensional embeddings obtained from the occurrence matrix of words used by participants in [3] to describe expressive piano performances.

References

[1] Chowdhury, S, Vall, A, Haunschmid, V, Widmer, G Towards Explainable Music Emotion Recognition: The Route via Mid-level Features. In Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR) 2019

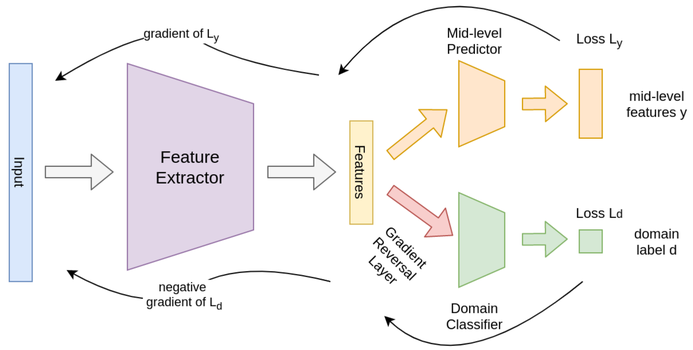
[2] Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." International conference on machine learning. PMLR, 2015.

[3] Cancino-Chacón, C, Peter, S, Chowdhury, S, Aljanaki, A, Widmer, G On the Characterization of Expressive Performance in Classical Music: First Results of the Con Espressione Game. In Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR) 2020.

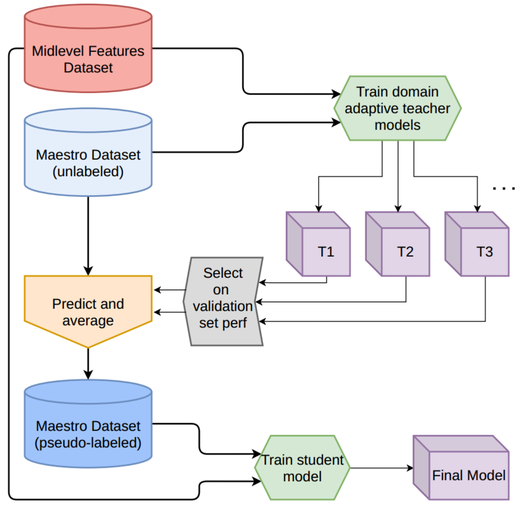
## GRADIENT-BASED UDA

We adopt the method of **gradient reversal** for unsupervised domain adaptation [2].

MAESTRO dataset is used as the target domain.




## TEACHER-STUDENT REFINEMENT



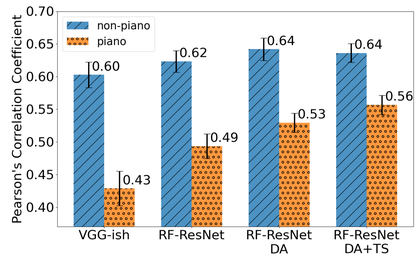

Only UDA results in some models not improving in performance on the target domain.

A refinement step helped overcome this problem.

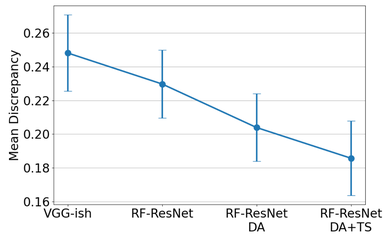
We selected the best-performing UDA models on the validation set, and used them as teacher models.

The student model trained using pseudo-labels from the teacher models performed better on average.

## RESULTS




Performance on piano and non-piano test sets




Mean discrepancy between piano and non-piano sets

## REGRESSION FOR EXPRESSIVE PERFORMANCE DESCRIPTORS

Evaluate whether our domain-adapted models can indeed predict better mid-level features for modelling the four dimensional expressive descriptor embeddings of the Con Espressione dataset [3].

| | Dim 1 | Dim 2 | Dim 3 | Dim 4 |
|---|---|---|---|---|
| VGG-ish | 0.35 | 0.10 | 0.22 | 0.32 |
| RF-ResNet | 0.36 | 0.07 | 0.28 | 0.33 |
| RF-ResNet DA | **0.40** | 0.09 | **0.29** | 0.32 |
| RF-ResNet DA+TS | 0.35 | **0.15** | **0.29** | **0.34** |

R2-score

We specifically look at dimension 1, the one that came out most clearly in the statistical analysis of the user responses characterized by a spectrum of descriptions like “hectic” and “agitated” to “calm” and “tender”.

| RF-ResNet | | RF-ResNet DA+TS | |
|---|---|---|---|
| Feature | $r$ | Feature | $r$ |
| articulation | 0.47 | melodiousness | − 0.39 |
| rhythmic complexity | 0.41 | articulation | 0.46 |
| | | rhythmic complexity | 0.41 |
| | | dissonance | 0.40 |

Pearson’s correlation (r) for mid-level features with the first description embedding dimension, with (right) and without (left) domain adaptation. Features with $p < 0.05$ and $|r| > 0.20$ are selected.

shreyan.chowdhury@jku.at

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 670035 ("Con Espressione")

European Research Council
Established by the European Commission