

DEEP RANKING: TRIPLET MATCHNET FOR MUSIC METRIC LEARNING

Rui Lu¹, Kailun Wu¹, Zhiyao Duan², Changshui Zhang¹

¹Department of Automation, Tsinghua University

²Department of Electrical and Computer Engineering, University of Rochester

March 8, 2017

Presentation at IEEE International Conference on Acoustics,
Speech and Signal Processing (ICASSP)

Applications of music metric learning

Classification

 <p>Composer: Brahms Orchestral, chamber, solo, and vocal music of Johannes Brahms</p>	 <p>Composer: Chopin The music of Frédéric François Chopin, the "poet of the piano"</p>	 <p>Composer: Dvorak The best of Antonin Dvorak</p>	 <p>Composer: Mahler The imaginative and unpredictable work of Gustav Mahler</p>
 <p>Composer: Mozart Performances of the work of Wolfgang Amadeus Mozart</p>	 <p>Composer: Tchaikovsky Scratch your itchy for Tchaikovsky</p>	 <p>Composer: Vivaldi Antonio Vivaldi, Italian Baroque composer of 'The Four Seasons'</p>	 <p>Composers: The "Three B's" A blend of Bach's best, boisterous Beethoven, and beautiful Brahms</p>



Recommendation

For You Recommendations Connect

FEBRUARY 23, 2017

Thursday's Playlists

- Best of Uptown Records
- Hip-Hop/R&B Hits: 1983
- Best of New Jack Swing, Vol. 1
- Behind the Boon and Tony Lewis

Recommended Albums

- James & Arthur
- ASAP Rocky
- JASO D'ARULO

Popular Songs

- Intro Fe
- It's Still Good
- OK OK
- My Away
- Call Call Rock
- I Need Love

Popular Albums

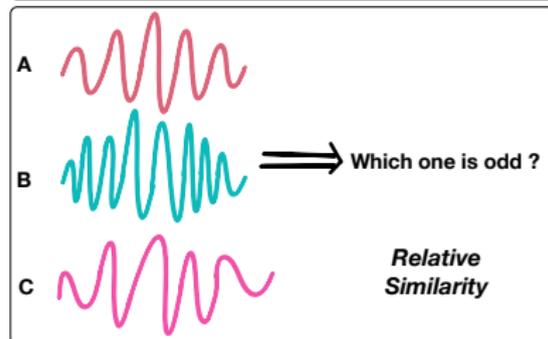
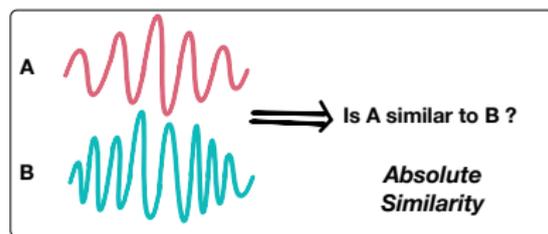
- Album cover of a woman's face

Music metric learning

Basic methods

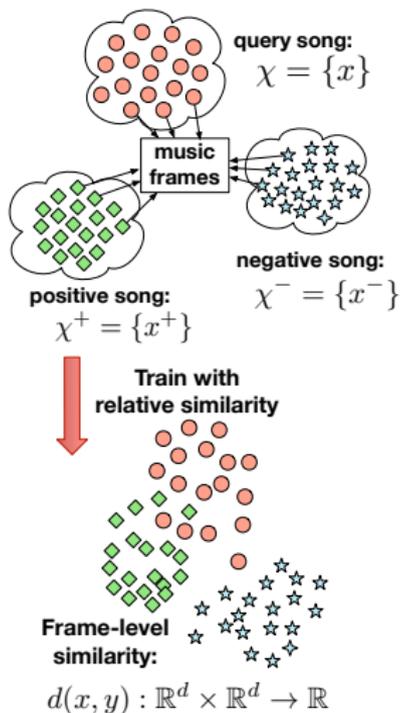
- Supervised methods
 - RITML: learns mahalanobis distance
 - MLR: learn to rank
 - SVM-based
 - ...
- Unsupervised methods
 - Mahalanobis distance
 - PCA
 - ...

Similarity

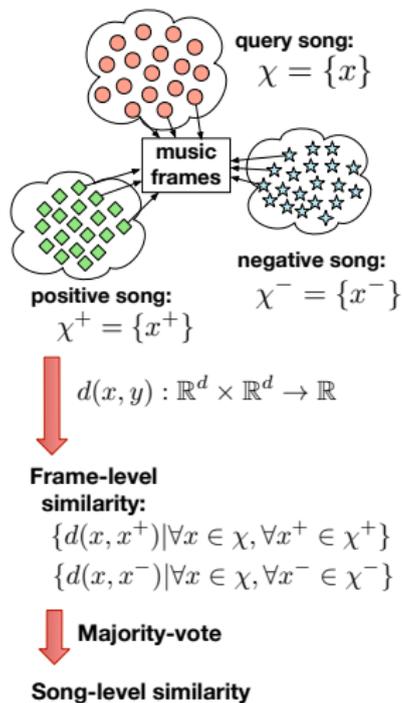


From frame-level to song-level

Training process



Testing process



Traditional and deep approaches

Traditional methods

- Handcrafted song-level features
- Linear projections

Deep learning approaches

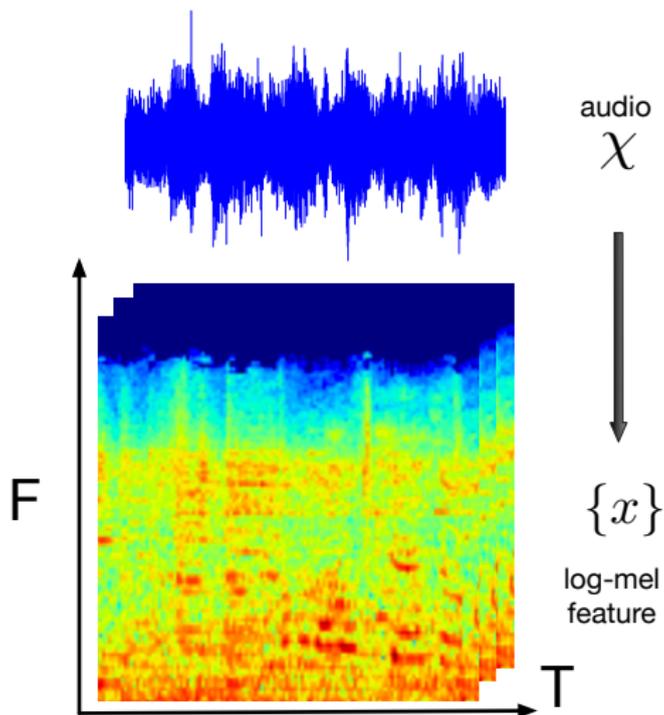
- Learn features automatically
- Highly nonlinear transformations
- Success in various domains
- None for music metric learning

Our approach

Use deep neural networks to learn frame-level relative similarities of music

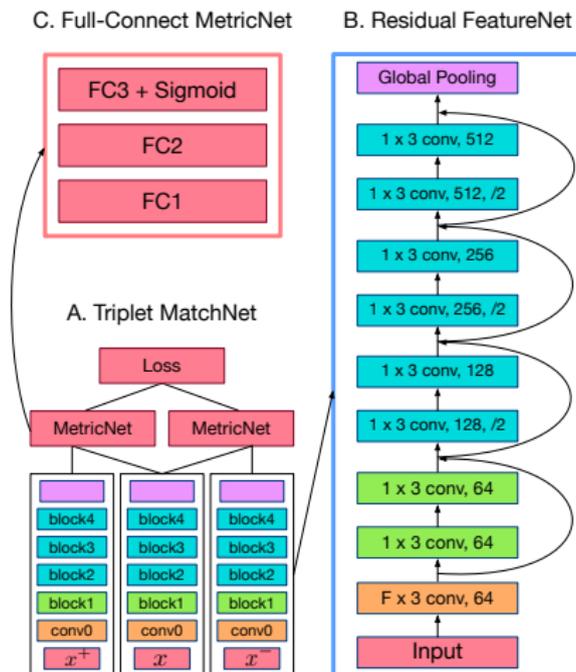
Method

Data preprocessing



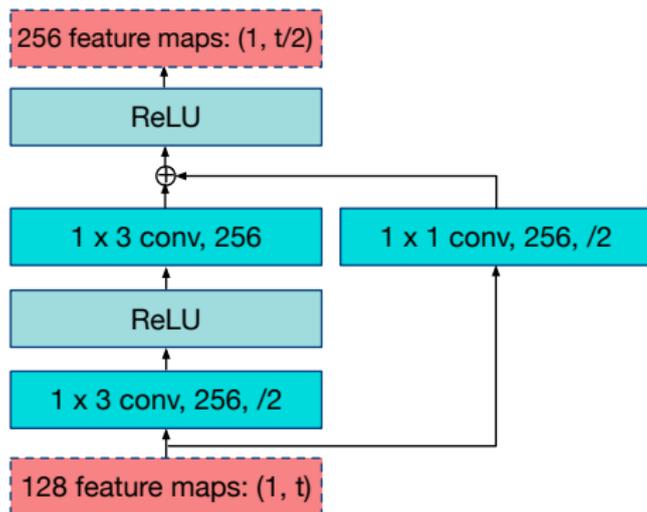
Triplet MatchNet

Network Structure



Residual block

Residual Structure



Advantages

- Easier to optimize
- Accuracy gain from deeper model^[1]
- Behave like ensembles of shallow networks^[2]

¹ Kaiming He et al, Deep residual learning for image recognition, CVPR 2016.

² Andreas Veit et al, Residual networks behave like ensembles of relatively shallow networks, NIPS 2016.

Final loss

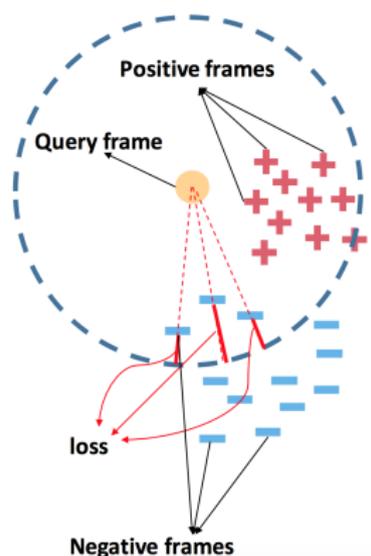
The final loss for training our Triplet MatchNet is:

$$\text{loss}(\mathcal{X}, \mathcal{X}^+, \mathcal{X}^-) = \frac{1}{|\{\mathcal{X}\}|} \sum_{x \in \{\mathcal{X}\}} (\psi(x) + \phi(x)).$$

Where $\psi(x)$ is the **rank-based loss**; $\phi(x)$ is the **contrastive loss**.

Rank-based loss

Illustration



Equation

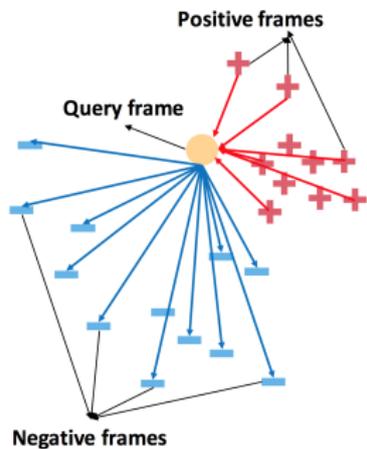
Rank-based loss

$$\psi(x) = \frac{1}{|\{x^-\}|} \sum_{x^- \in \{x^-\}} \max\{0, d_{max}^+ - f(x, x^-)\}$$

where $d_{max}^+ = \max_{x^+ \in \{x^+\}} f(x, x^+)$;
 $f(x, y)$ is the proposed network's final output

Contrastive loss

Illustration



Equation

Contrastive loss

$$\phi(x) = -\frac{\sum_{x^+} \sum_{x^-} [\log(1 - d^+) + \log(d^-)]}{|\{x^+\}| |\{x^-\}|}$$

where $d^+ = f(x, x^+)$; $d^- = f(x, x^-)$

Experiments

Dataset and Evaluation

MagnaTagATune

- Relative similarity
- 860 triplets like (χ, χ^+, χ^-)
- 993 unique songs
- Each song with 29 seconds

Evaluation

- Constraints Fulfillment Rate
 - Portion of triplets that preserve partial order relationships
- 10-cross validation
- Comparison methods
 - RITML^[1]
 - MLR^[2], RMLR^[3]
 - SVM, Euclidean

¹ Daniel Wolff et al, Comparative music similarity modelling using transfer learning across user groups, ISMIR 2015.

² Brian McFee et al, Metric learning to rank, ICML 2010.

³ Daryl KH Lim et al, Robust structural metric learning, ICML 2013

Constraints Fulfillment Rate comparison

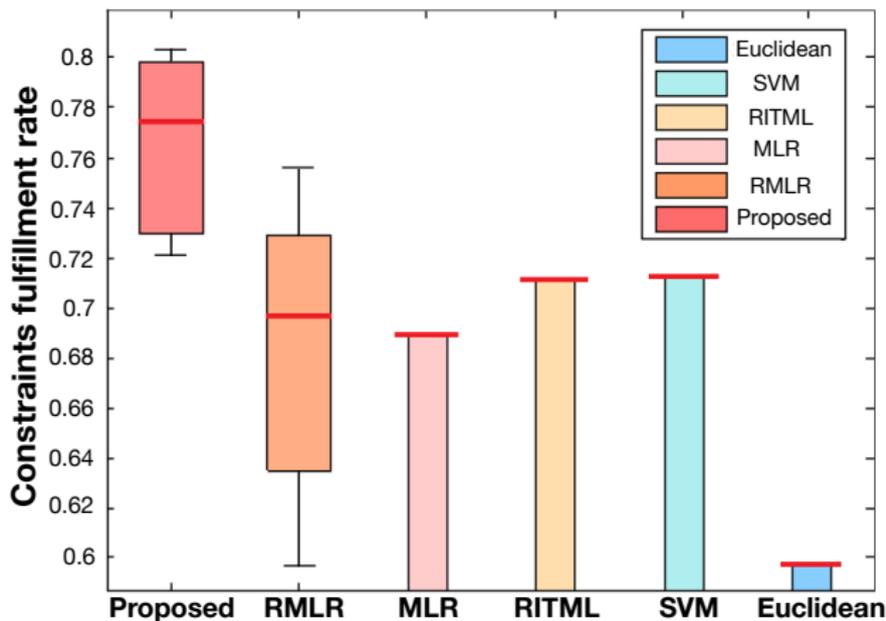


Figure: Constraints Fulfillment Rate by 10-fold cross validation.

Generalization Capability

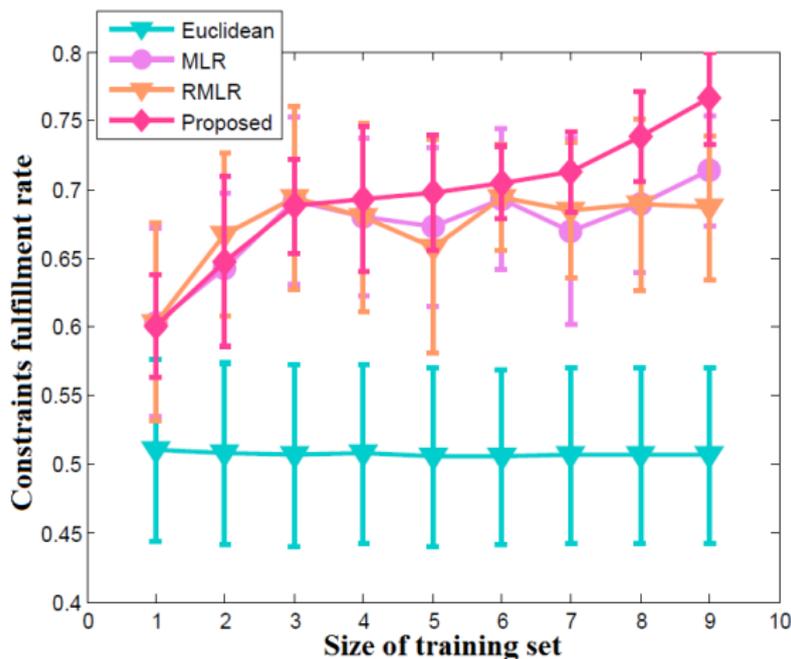


Figure: Generalization capability by 10-fold cross validation.

Better features extracted by Triplet MatchNet

Method	HandCrafted	PCA	Proposed
RMLR	-	65.9 ± 8.3	71.2 ± 7.2
MLR	68.9	61.7 ± 10.5	71.7 ± 6.9
Euclidean	59.8	50.7 ± 6.3	70.6 ± 3.8

Table: Constraints Fulfillment Rate of three baselines working with different features.

Thank you !