"AI for Visual and Musical Arts: Recommendation Systems, Intelligent User Interfaces and Generative Models"

Denis Parra
Associate Professor, CS Department
School of Engineering
Pontificia Universidad Católica de Chile & IMFD

Data Science & Computational Social Science Seminar, University of Michigan, 2020

Biografía

- Associate Prof. PUC Chile, member of: AI Lab, HAIVis UC y CreativAI Lab UC.
- Researcher at Millenium Institute Foundational Research on Data (10-year initiative, 5 institutions)
- Research on: Recommendation Systems, Applied AI (Healthcare, Creative AI), and Visualization

Topics in this talk

Topic 1:

Brief intro to Recommendation Systems
+
Projects developed at PUC Chile

Topic 2:

Brief intro to Deep Generative Models
+
Projects developed at PUC Chile

TOC

- Topic 1 – Recommender Systems
  - Explorative RecSys: Moodplay, MNBA
  - Art Recommendation using Deep Neural Networks: VisRank y CuratorNet
- Topic 2 – Generative Models
  - Visual Generation: Transfer style with AdaIN …then… styleGAN2 for the win!
  - Music Generation: latent chords (TimbreNet) and MIDIGAN
- Projects under development
- Ideas to discuss

Topic 1:

Brief intro to Recommendation Systems

+

Projects developed at PUC Chile

Why are RecSys Important ?

- Recommender Systems (RecSys) have become important due to the large number of applications where people need to make decisions among many items

NETFLIX Spotify YouTube

RecSys Definition

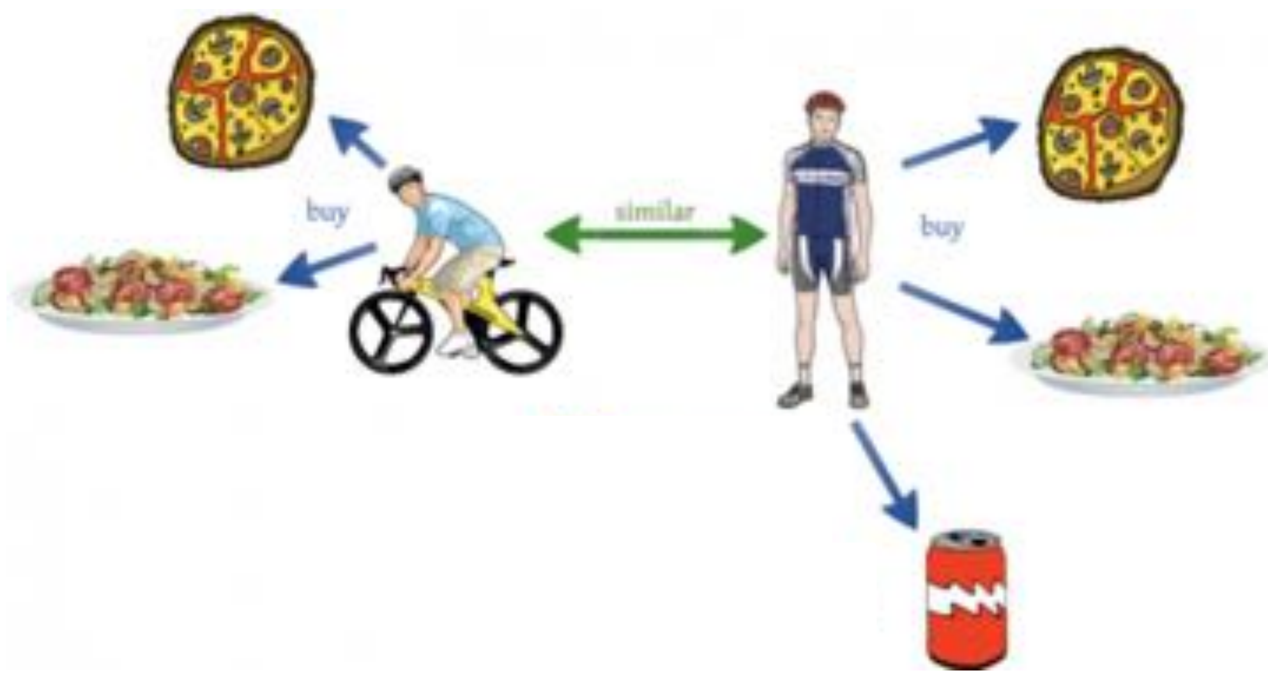
Systems which help (groups of) people find relevant items in an overloaded information space.

(MacNee et al. 2006)

¿ Which are the most important methods ?

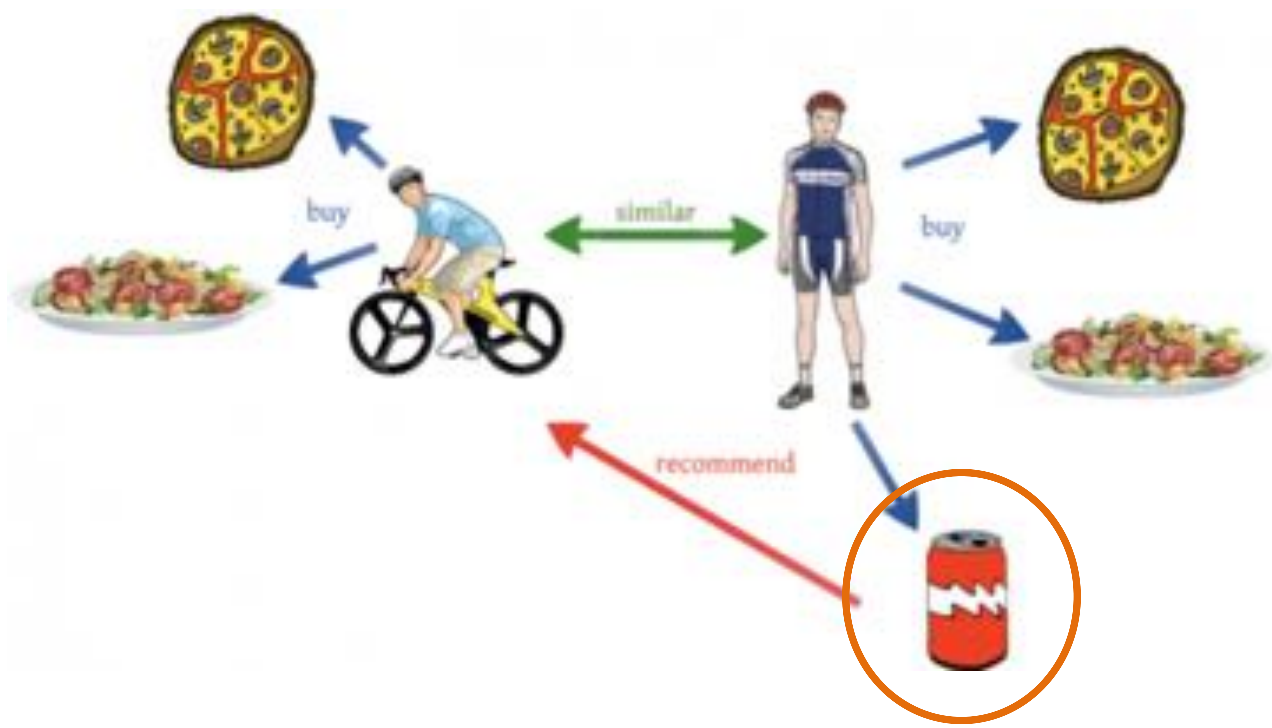
Collaborative Filtering

- Identify like-minded people …



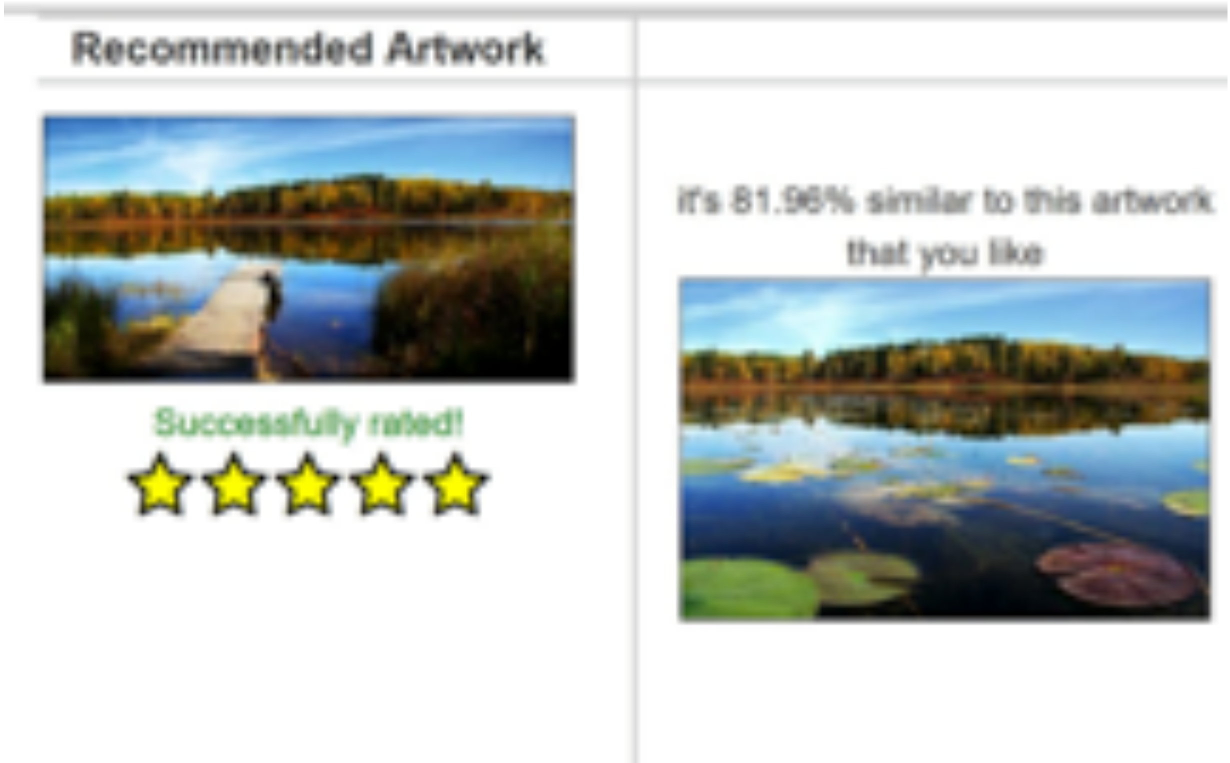
Filtrado colaborativo

- Identify like-minded people …then recommend items not consumed yet



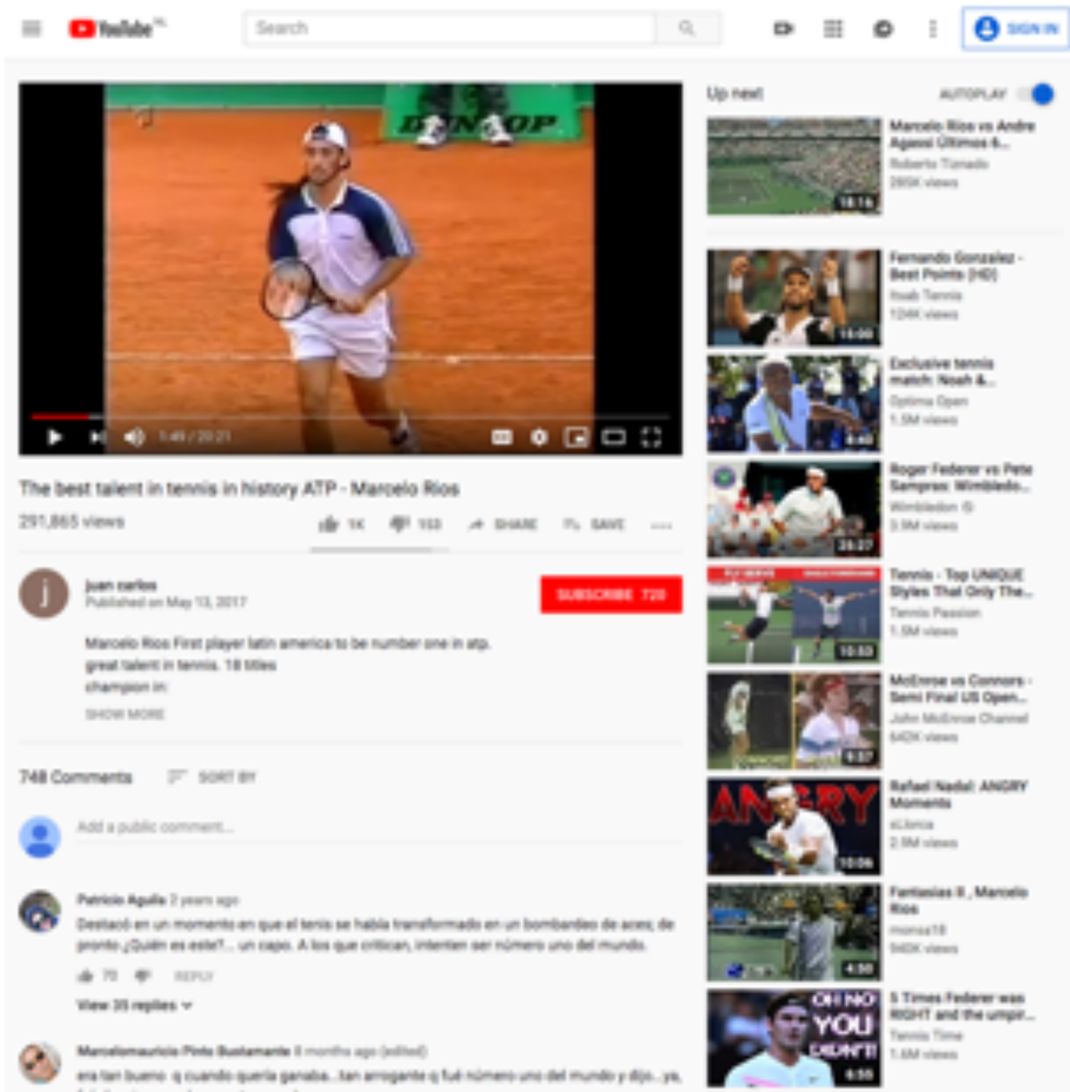
Content-based filtering

- Use content of items to identify similarity, and then make recommendations
- This approach allows to deal with cold-start and new item problems.



¿ How do recommendations are usually presented ?

In YouTube

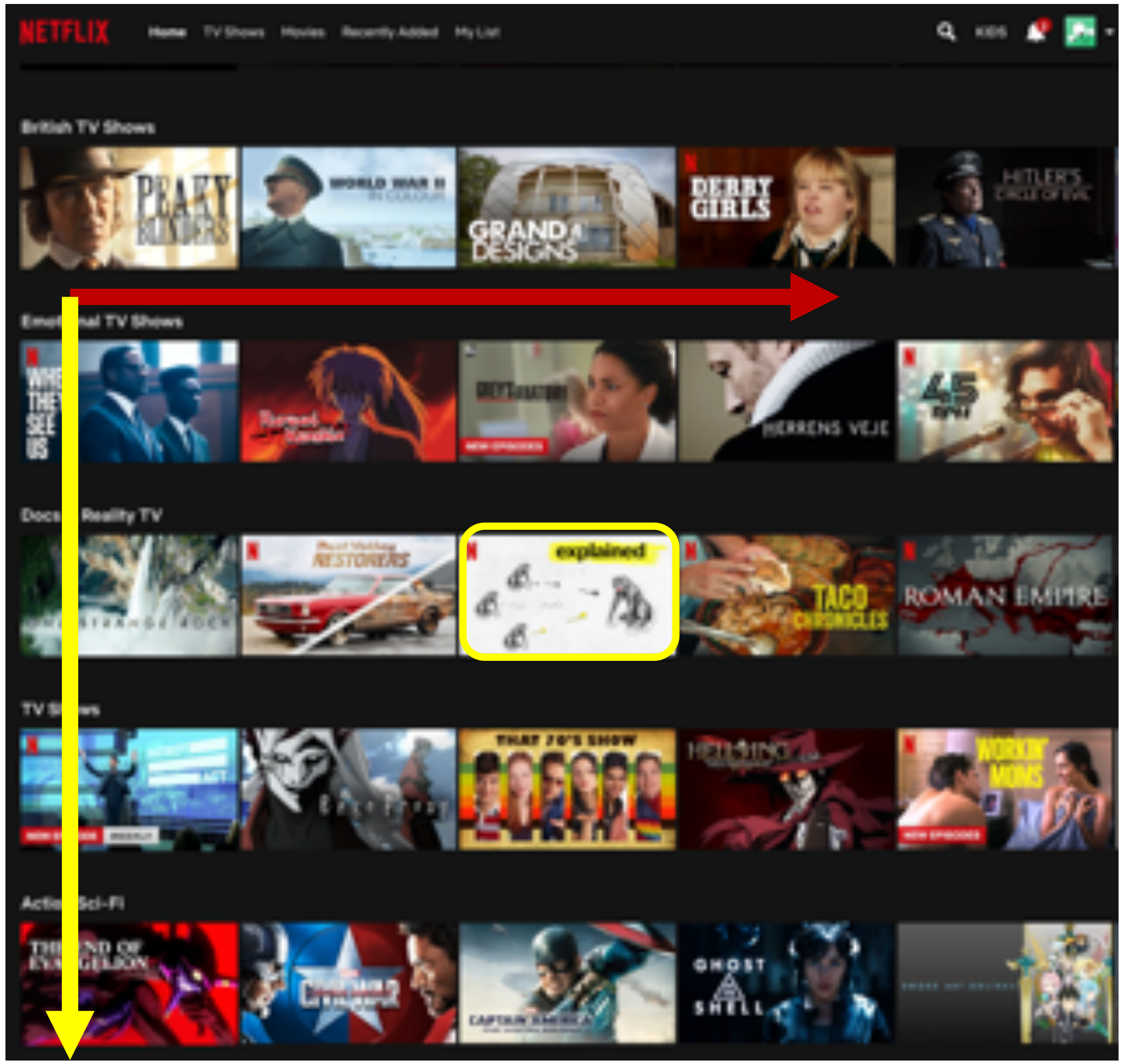


- Ranked list of recommended videos

YouTube - Personalization

- If I open the same video in a different country, I get different recommendations.
- Different levels of contextualization and personalization.

In Netflix



- Lists of movies and TV series
- Ranking: lists of movies and items within the lists
- Cover image to present the movie (reinforcement learning)

Topic 1:

Brief intro to Recommendation Systems

+

Projects developed at PUC Chile

Research and Development at PUC

- Research and Systems developed at PUC (2016 -)
  - Moodplay (demo) http://moodplay.pythonanywhere.com/
  - MNBA Surdoc http://niebla.ing.puc.cl/surdoc/
  - Neural Networks for visual art representation (VisRank)
  - CuratorNet (demo)

Moodplay

- Goals: Recommend musical artists
- Use exploration rather than ranked lists (give users control – Verbert et al. (2013))
- Represent artists by the emotion that people feel when they listen to them, GEMS model
- http://moodplay.pythonanywhere.com

I. Andjelkovic

J. O’Donovan

R. Herrera

Publications

- Andjelkovic, I., Parra, D., & O'Donovan, J. (2019). Moodplay: Interactive music recommendation based on Artists' mood similarity. *International Journal of Human-Computer Studies*, *121*, 142-159.
- Andjelkovic, I., Parra, D., & O'Donovan, J. (2016). Moodplay: Interactive mood-based music discovery and recommendation. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (pp. 275-279).
- Verbert, K., Parra, D., et al. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. 2013. p. 351-362.

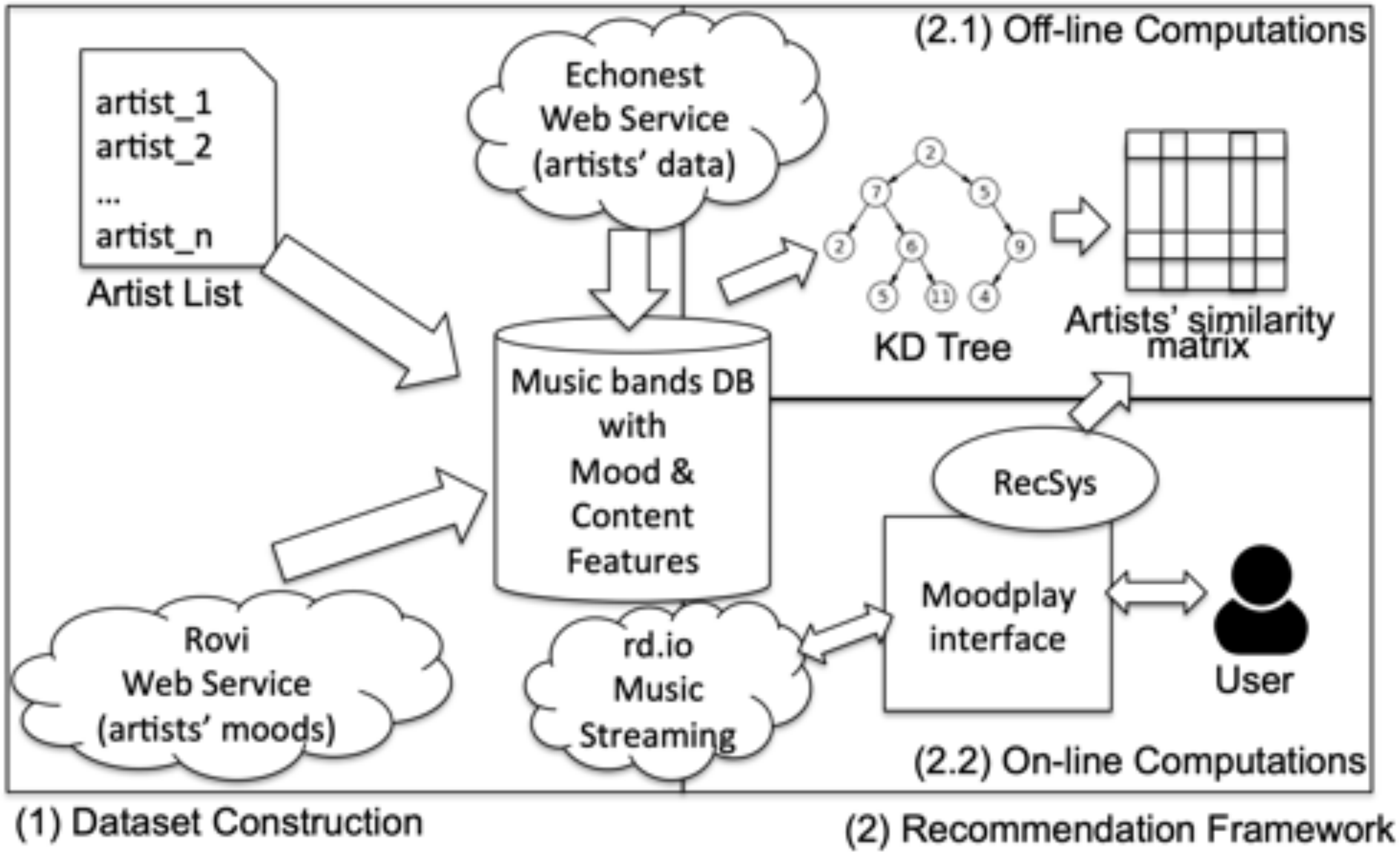
GEMS Model

- Rather than traditional [Valence, Arousal, Dominance] emotion model, we used music-specific emotion model GEMS

Category	Sub-category	No. of moods	Example moods
Sublimity	Tenderness	24	Delicate, romantic, sweet
	Peacefulness	22	Pastoral, relaxed, soothing
	Wonder	24	Happy, light, springlike
	Nostalgic	9	Dreamy, rustic, yearning
	Transcendence	10	Atmospheric, spiritual, uplifting
Vitality	Power	29	Ambitious, fierce, pulsing, intense
	Joyful activation	32	Animated, fun, playful, exciting
Unease	Tension	32	Nervous, harsh, rowdy, rebellious
	Sadness	18	Austere, bittersweet, gloomy, tragic
	Fear *	10	Spooky, nihilistic, ominous
	Lethargy *	8	Languid, druggy, hypnotic
	Repulsiveness *	10	Greasy, sleazy, trashy, irreverent
Other *	Stylistic *	19	Graceful, slick, elegant, elaborate
	Cerebral *	12	Detached, street-smart, ironic
	Mechanical *	7	Crunchy, complex, knotty

http://www.zentnerlab.com/psychological-tests/geneva-emotional-music-scales

Software Architecture



Recommender System Hybrid Model

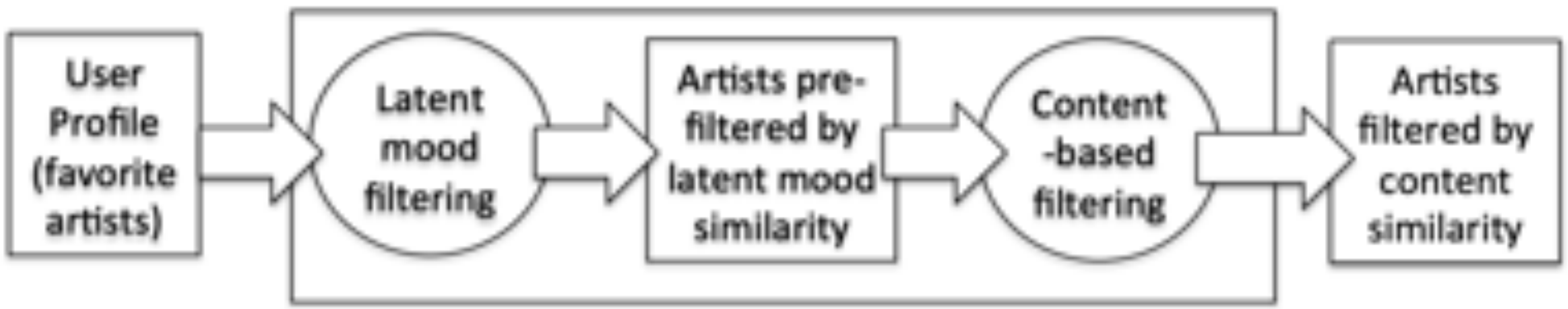
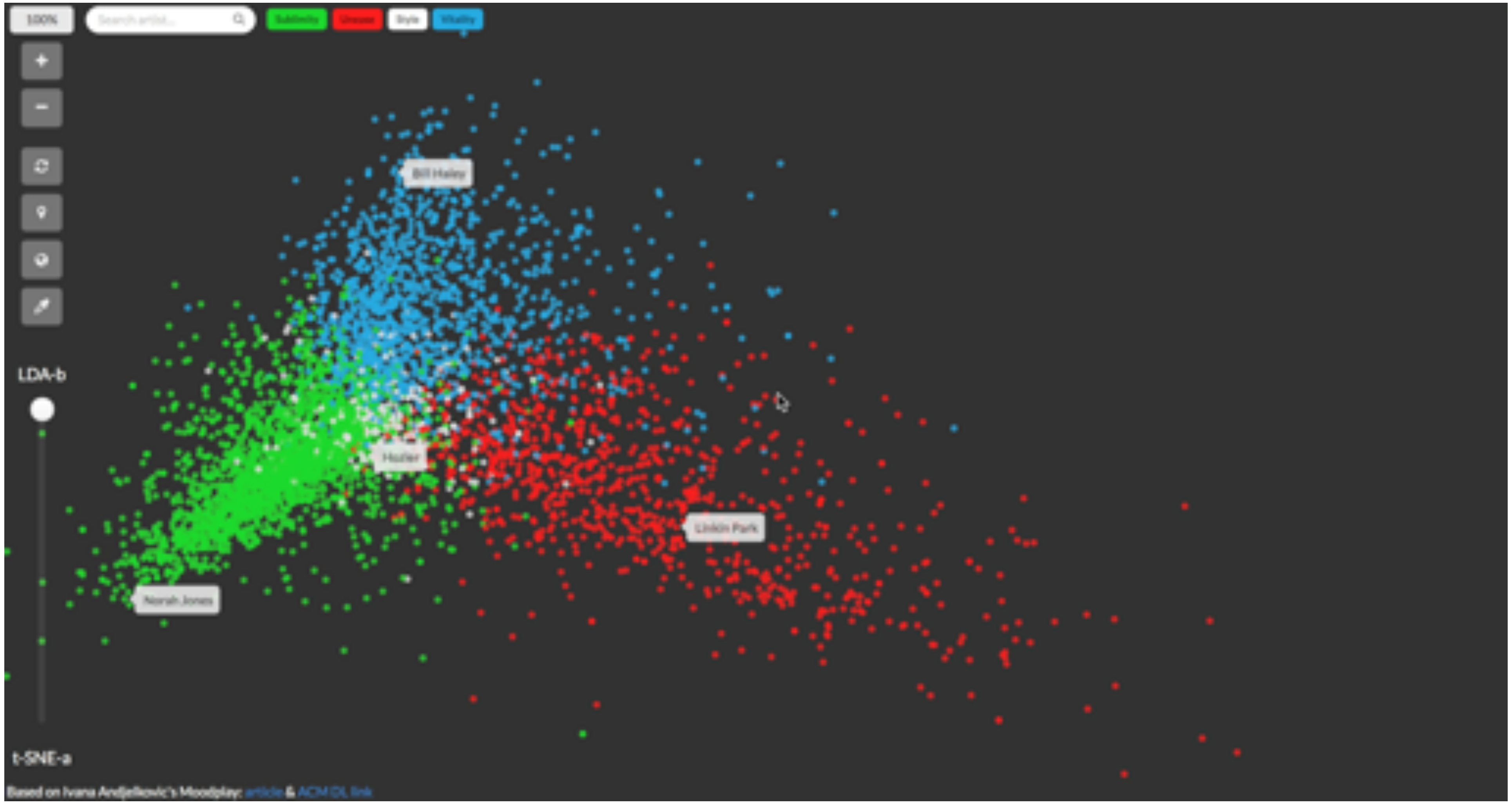


Figure 4: Schematic representing our hybrid cascading recommender which pre-filters based on mood similarity and then post-filters based on content similarity.

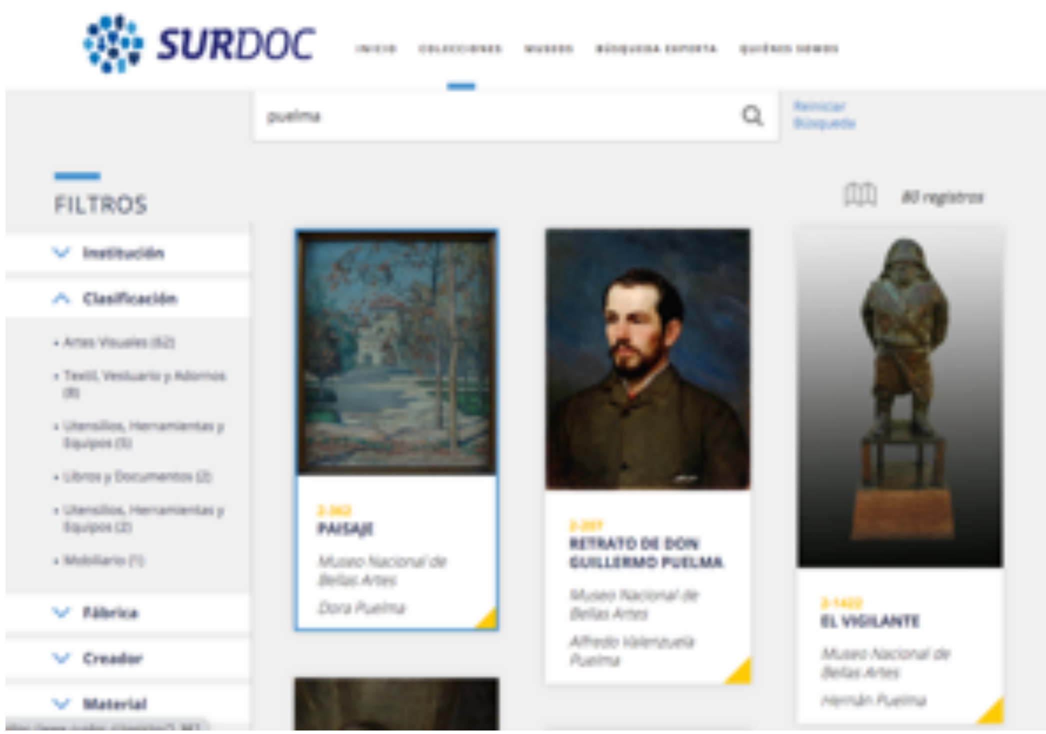
Moodplay Demo, version 2



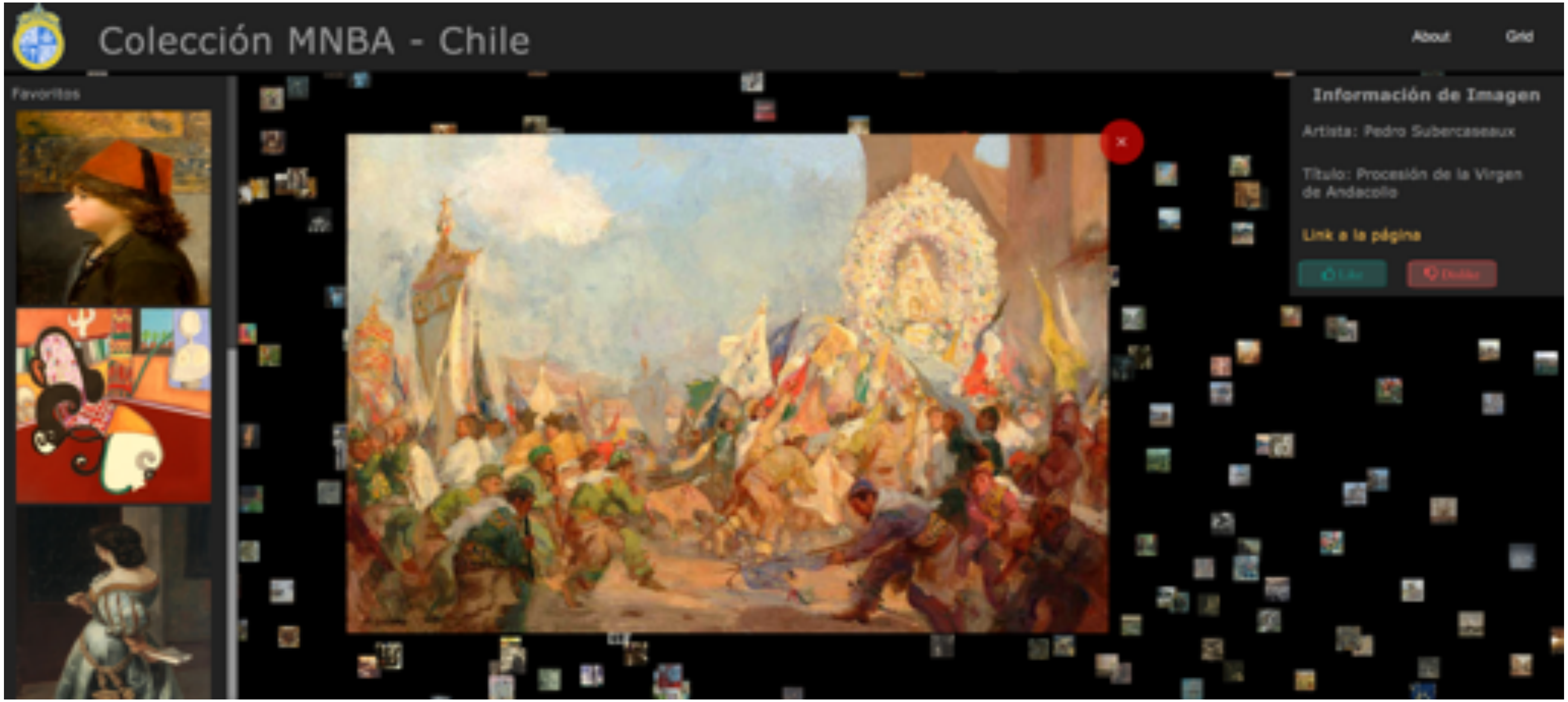
Version1 Demo : https://www.youtube.com/watch?v=vH9q5ku8ocM&feature=youtu.be

MNBA Explorer - Surdoc

- Goal: facilitate exploration and discovery of paintings at MNBA
- Current Museum portal centered on search/filtering



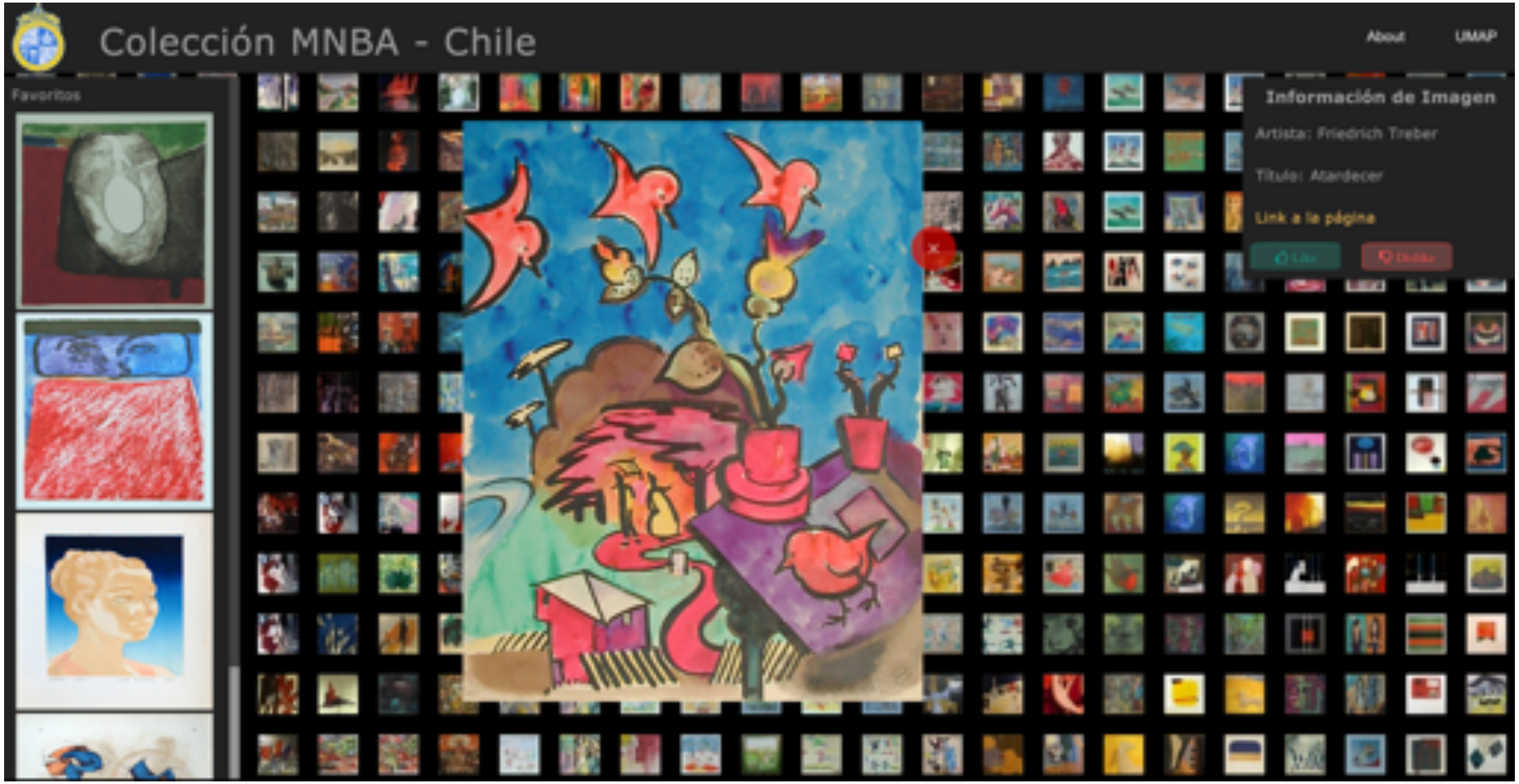
MNBA Explorer – Surdoc



R. Schilling

https://niebla.ing.puc.cl/surdoc/

MNBA Explorer – Surdoc



R. Schilling

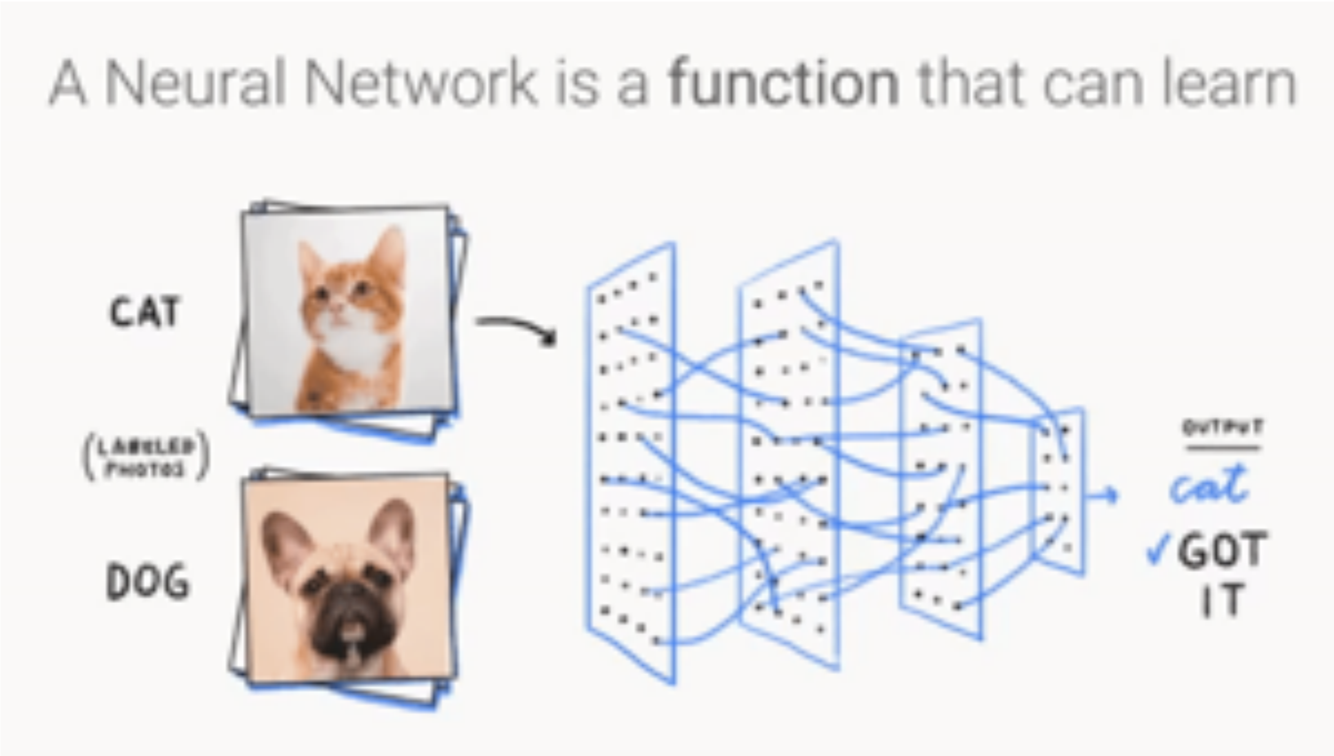
https://niebla.ing.puc.cl/surdoc/

How do we represent images ?

- Traditionally, Manually engineered features:
  - Local binary patterns (LBP)
  - SIFT
  - Color histograms
  - Contrast
  - Luminosity
  - Etc.

Neural Networks

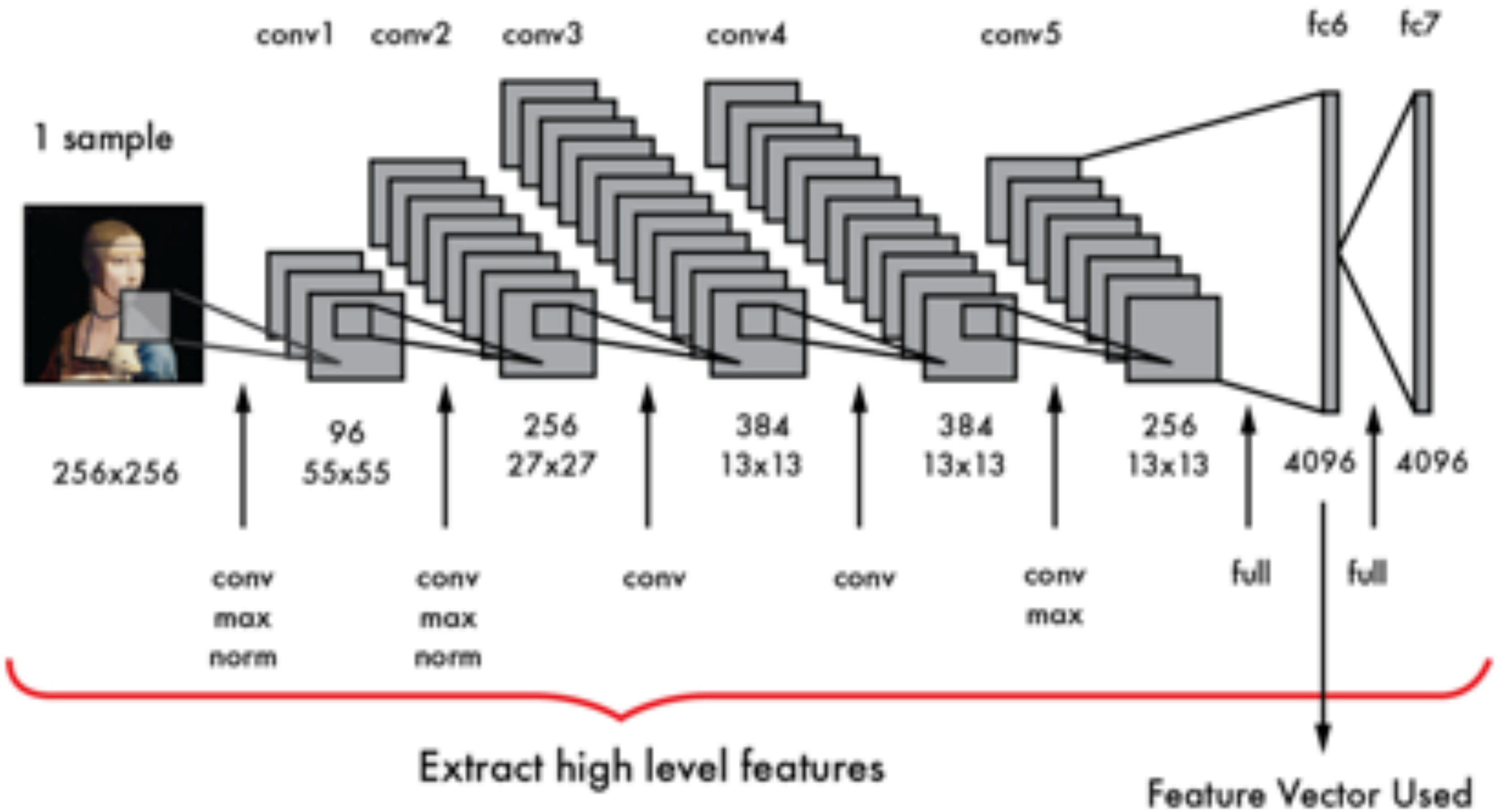
- A type of model that learns a task from data, amazing at finding patterns and learning features automatically



[Image from https://mc.ai/aisaturdaylagos-the-torch-panther/](https://mc.ai/aisaturdaylagos-the-torch-panther/)

Why do we use them?

- Great for learning features, which can later be used for other tasks (transfer learning)



Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks.

VisRank & CuratorNet

- Publications:
  - Messina, P., Cartagena, M., Cerda, P., del Rio, F., & Parra, D. (2020). CuratorNet: Visually-aware Recommendation of Art Images. arXiv preprint arXiv:2009.04426. *Proceedings of ComplexRec-ImpactRS workshop, co-located at RecSys 2020*

  - Messina, P., Dominguez, V., Parra, D., Trattner, C., & Soto, A. (2019). Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features. *User Modeling and User-Adapted Interaction*, *29*(2), 251-290.

VisRank team

Vicente Domínguez

Pablo Messina

Ivania Donoso-Guzmán

Christoph Trattner

Alvaro Soto

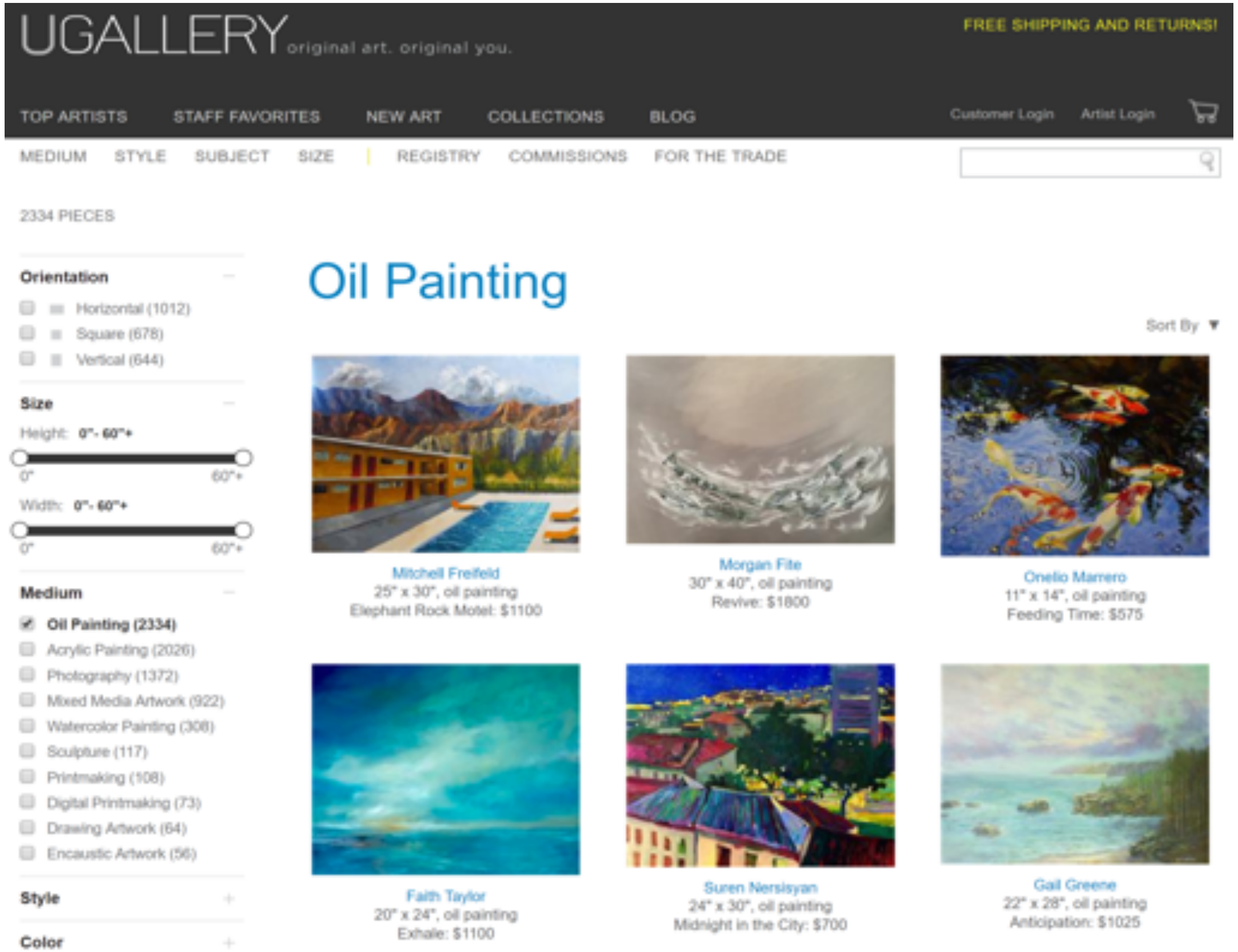
Manuel Cartagena

Denis Parra

Domingo Mery

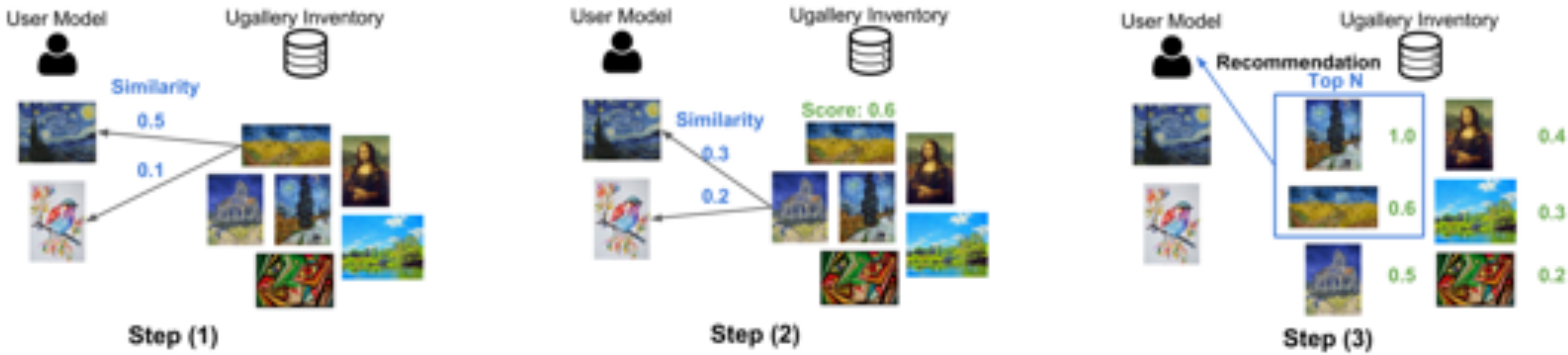
Context: Ugallery

- Online art store
- Focuses specially in emergent artists: to help them sell their paintings
- Our partners in this project



Content-based Recommendation

- User Transactions on mostly one-of-a-kind paintings, difficult to do collaborative filtering



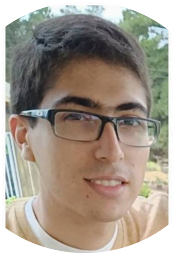
Transfer Learning and fine-tuning

- del Rio, F., Messina, P., Dominguez, V., & Parra, D. (2018). Do Better ImageNet Models Transfer Better... for Image Recommendation?. *arXiv preprint arXiv:1807.09870.*
- ResNet was the best for transfer learning (with and w/o finetuning)

Table 1: Results of different pre-trained embeddings at the artwork image recommendation task to the left (R:Recall, P:Precision), and their performance at the ILSVRC Challenge trained on ImageNet dataset (Acc: Accuracy). The top methods in both tasks do not correlate.

CNN	Artwork Image Recommendation				ILSVRC-2012-CLS	
	R@20	P@20	MRR@20	nDCG@20	Top-1 Acc. (%)	Top-5 Acc. (%)
ResNet50	**.1632**	**.0141**	**.0979**	**.1253**	75.2	92.2
VGG19	**.1398**	**.0124**	**.0750**	**.1008**	71.1	89.8
NASNet Large	.1379	.0120	.0743	.0998	**82.7**	**96.2**
InceptionV3	.1332	.0125	.0744	.1007	78.0	93.9
InceptionResNetV2	.1302	.0117	.0692	.0936	**80.4**	**95.3**
Random	.0172	.0013	.0051	.0093	-	-

CuratorNet team



Pablo Messina

Manuel Cartagena

Felipe del Río

Patricio Cerda

Denis Parra

CuratorNet

- Inspired by both VBPR (2015) and YouTube Recsys (2016)

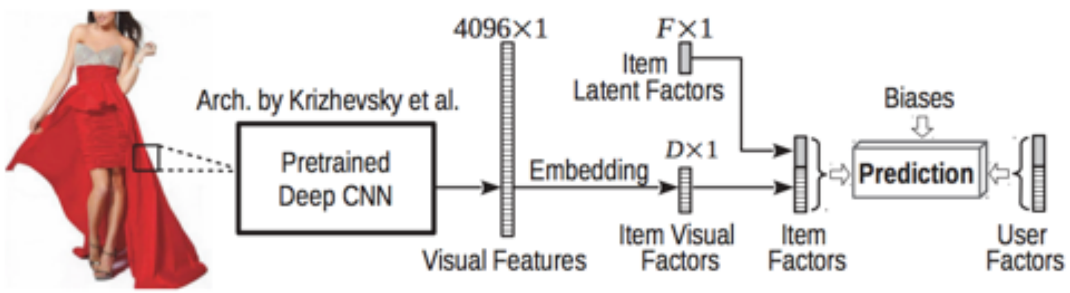
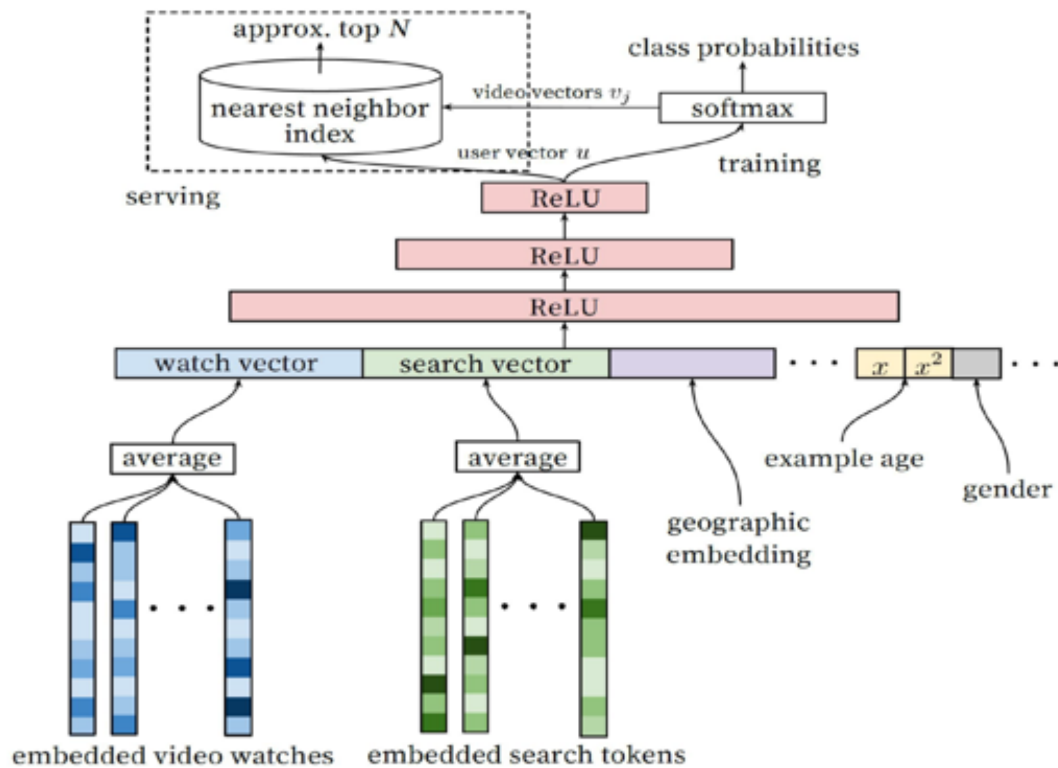


Figure 1: Diagram of our preference predictor. Rating dimensions consist of visual factors and latent (non-visual) factors. Inner products between users and item factors model the compatibility between users and items.

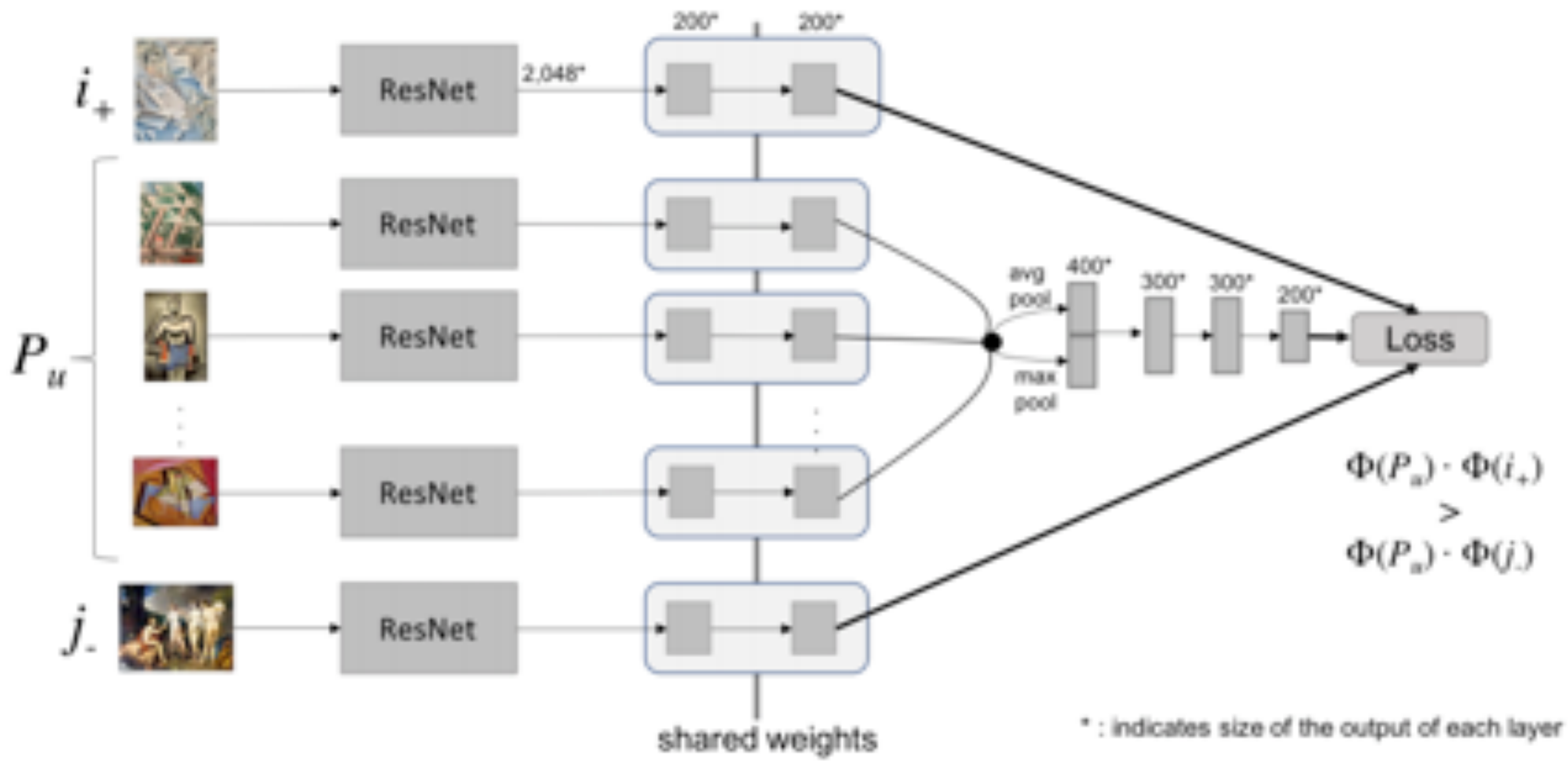


He, R., & McAuley, J. (2016). VBPR: visual Bayesian Personalized Ranking from implicit feedback. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (pp. 144-150).

Covington, P., Adams, J., & Sargin, E. (2016, September). Deep neural networks for youtube recommendations. In Proceedings of the 10th ACM conference on recommender systems (pp. 191-198).

CuratorNet

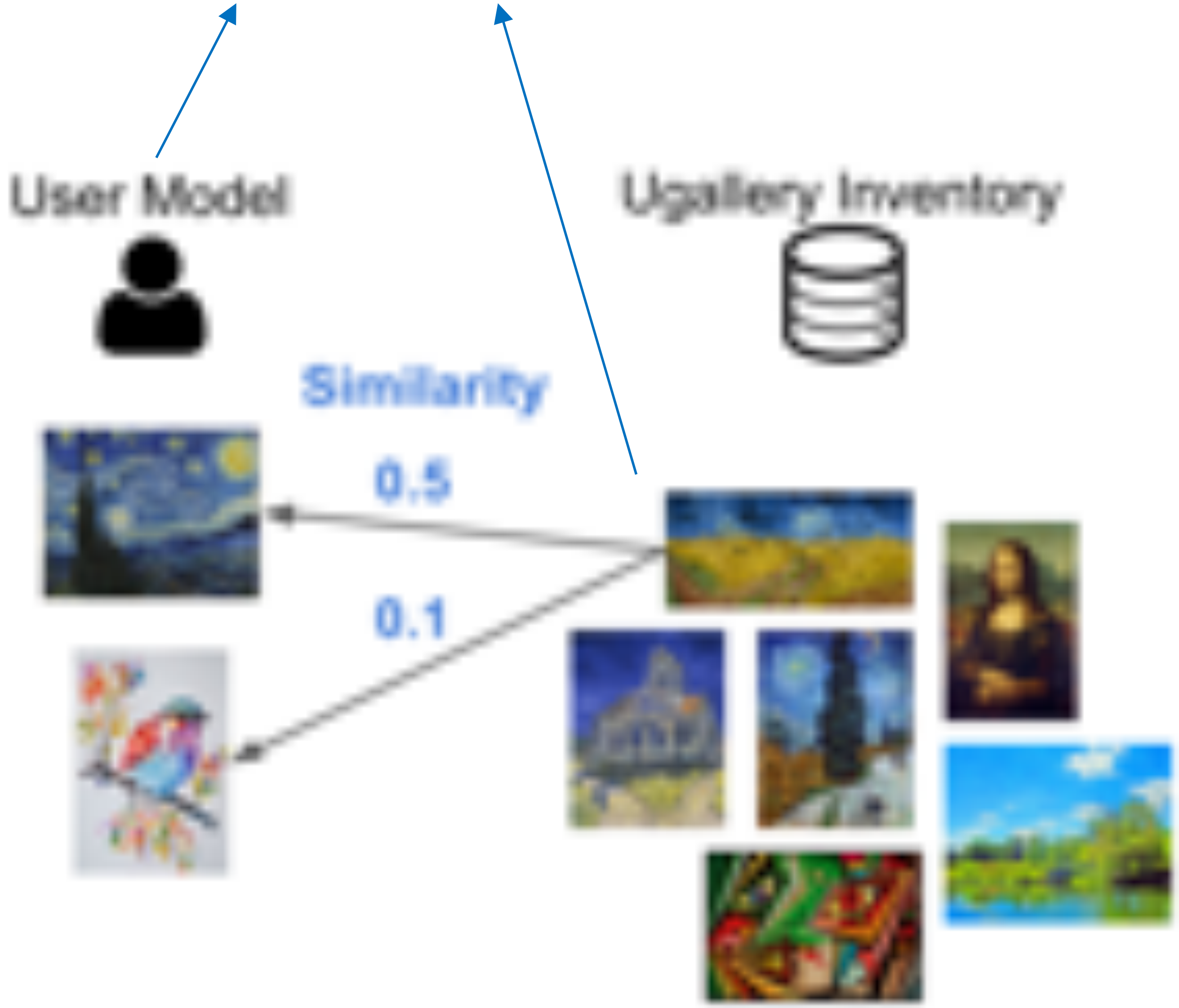
- Using BPR framework, we learn from triples (P_u, i+, j-)



Messina, P., Cartagena, M., Cerda, P., del Rio, F., & Parra, D. (2020). CuratorNet: Visually-aware Recommendation of Art Images. Proceedings of ComplexRec-ImpactRS workshop, co-located at RecSys 2020

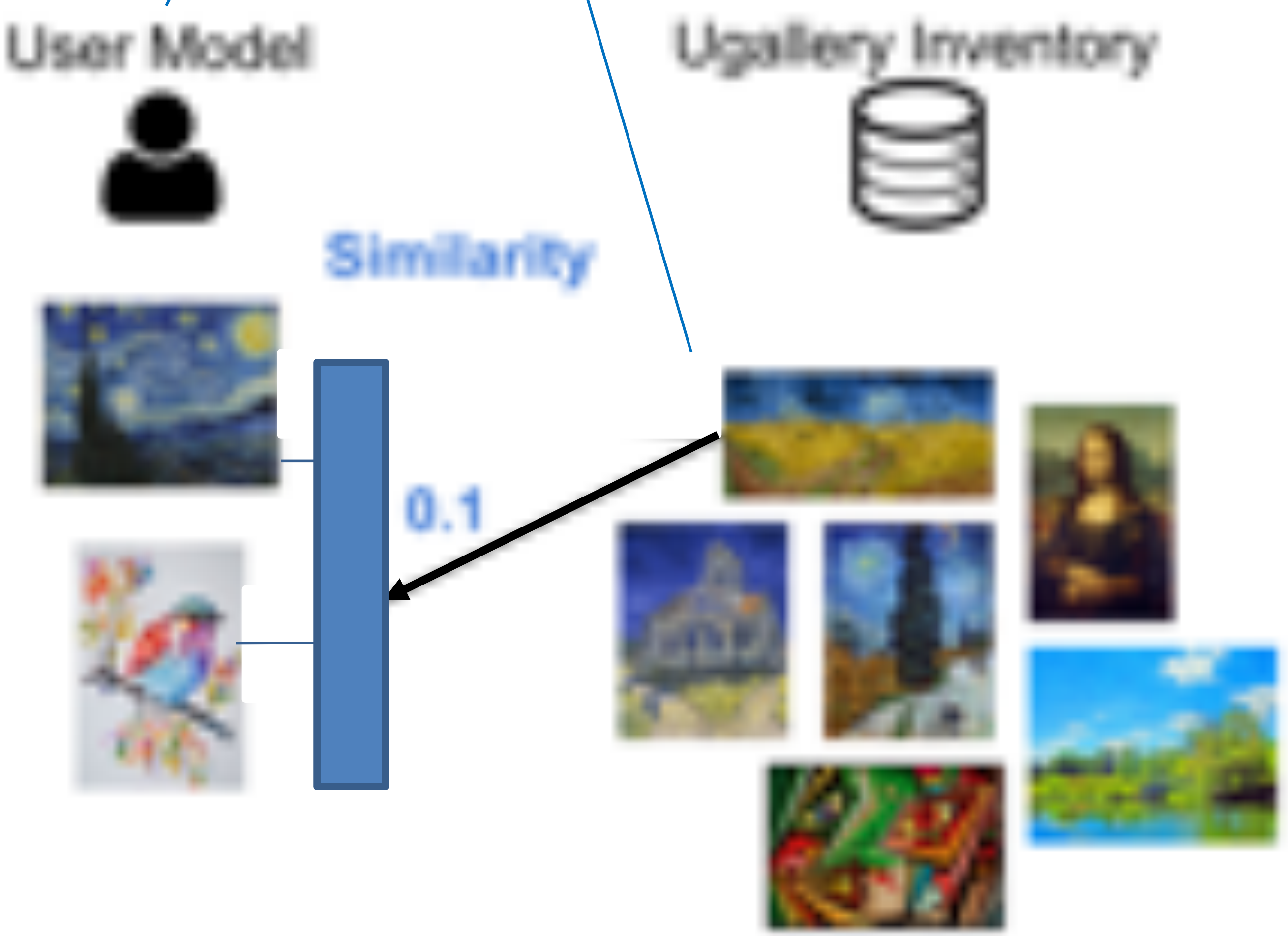
VisRank (baseline)

$s(u,i) \;= score(\,ContentBasedProfile(u),\; Content(i)\,)$



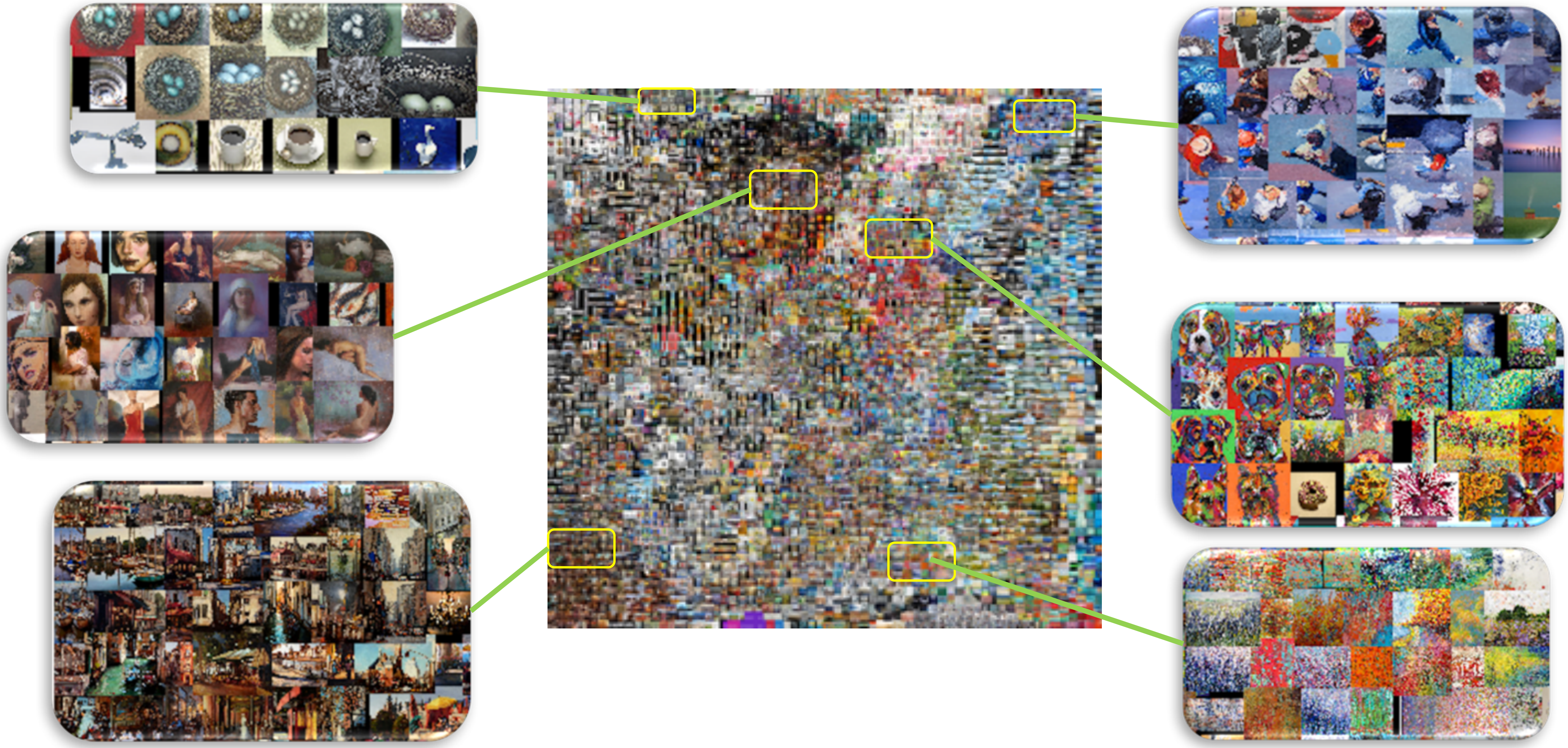
CuratorNet

$s(u,i) \;= score(\,ContentBasedProfile(u),\; Content(i)\,)$



User Model representation by CuratorNet

Guidelines for negative sampling



Results: VisRank, VBPR, CuratorNet

Method	λ (L2 Reg.)	AUC	R@20	P@20	nDCG@20	R@100	P@100	nDCG@100
Oracle	–	**1.0000**	**1.0000**	**.0655**	**1.0000**	**1.0000**	**.0131**	**1.0000**
CuratorNet	.0001	**.7204**	**.1683**	**.0106**	**.0966**	**.3200**	**.0040**	**.1246**
CuratorNet	.001	**.7177**	**.1566**	**.0094**	**.0895**	**.2937**	**.0037**	**.1160**
VisRank	–	**.7151**	**.1521**	**.0093**	**.0956**	**.2765**	**.0034**	**.1195**
CuratorNet	0	**.7131**	**.1689**	**.0100**	**.0977**	**.3048**	**.0038**	**.1239**
CuratorNet	.01	.7125	.1235	.0075	.0635	.2548	.0032	.0904
VBPR	.0001	.6641	.1368	.0081	.0728	.2399	.0030	.0923
VBPR	0	.6543	.1287	.0078	.0670	.2077	.0026	.0829
VBPR	.001	.6410	.0830	.0047	.0387	.1948	.0024	.0620
VBPR	.01	.5489	.0101	.0005	.0039	.0506	.0006	.0118
Random	–	.4973	.0103	.0006	.0041	.0322	.0005	.0098

Guidelines for negative sampling

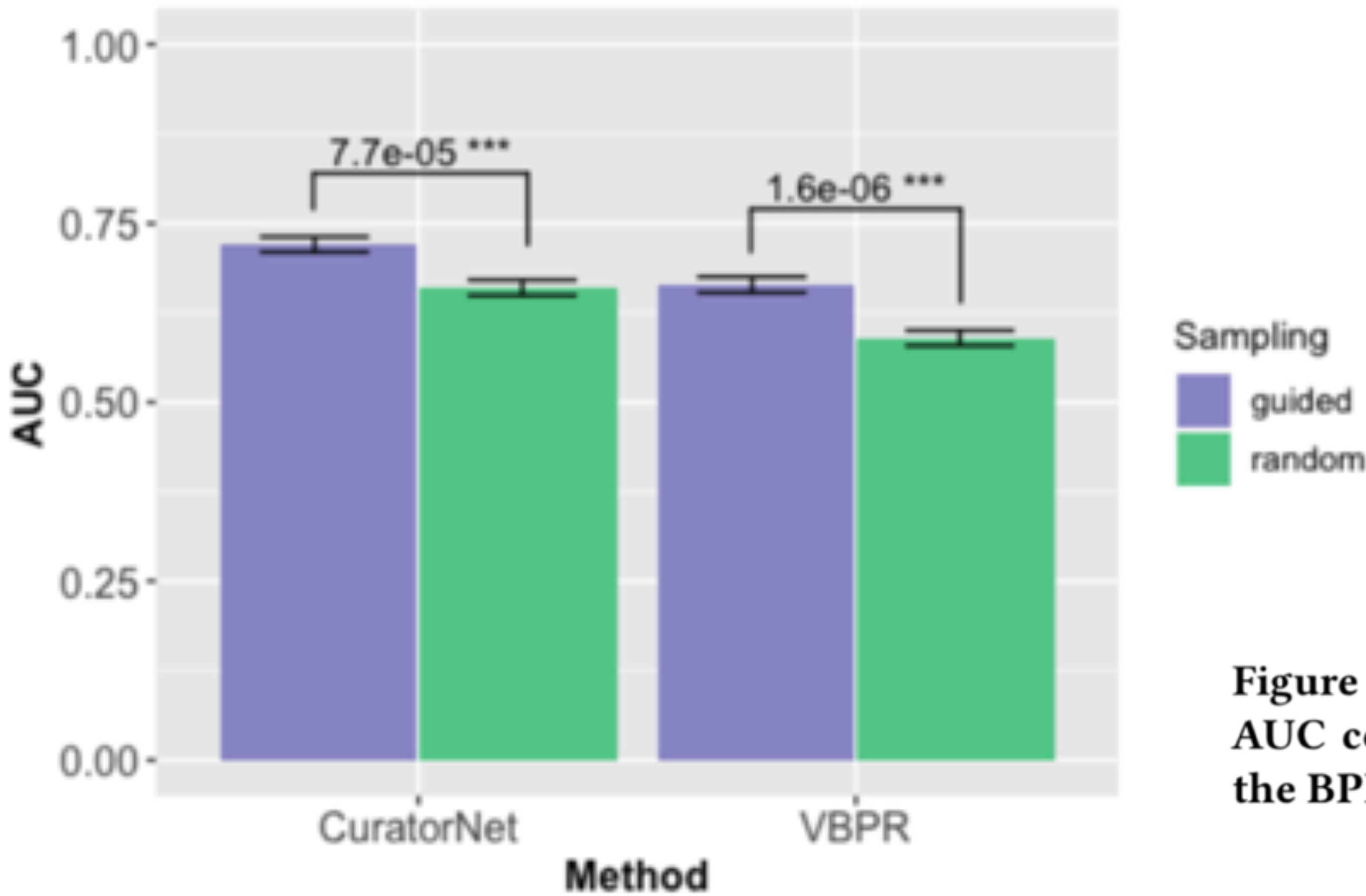


Figure 4: The sampling guidelines had a positive effect on AUC compared to random negative sampling for building the BPR training set.

Conclusion

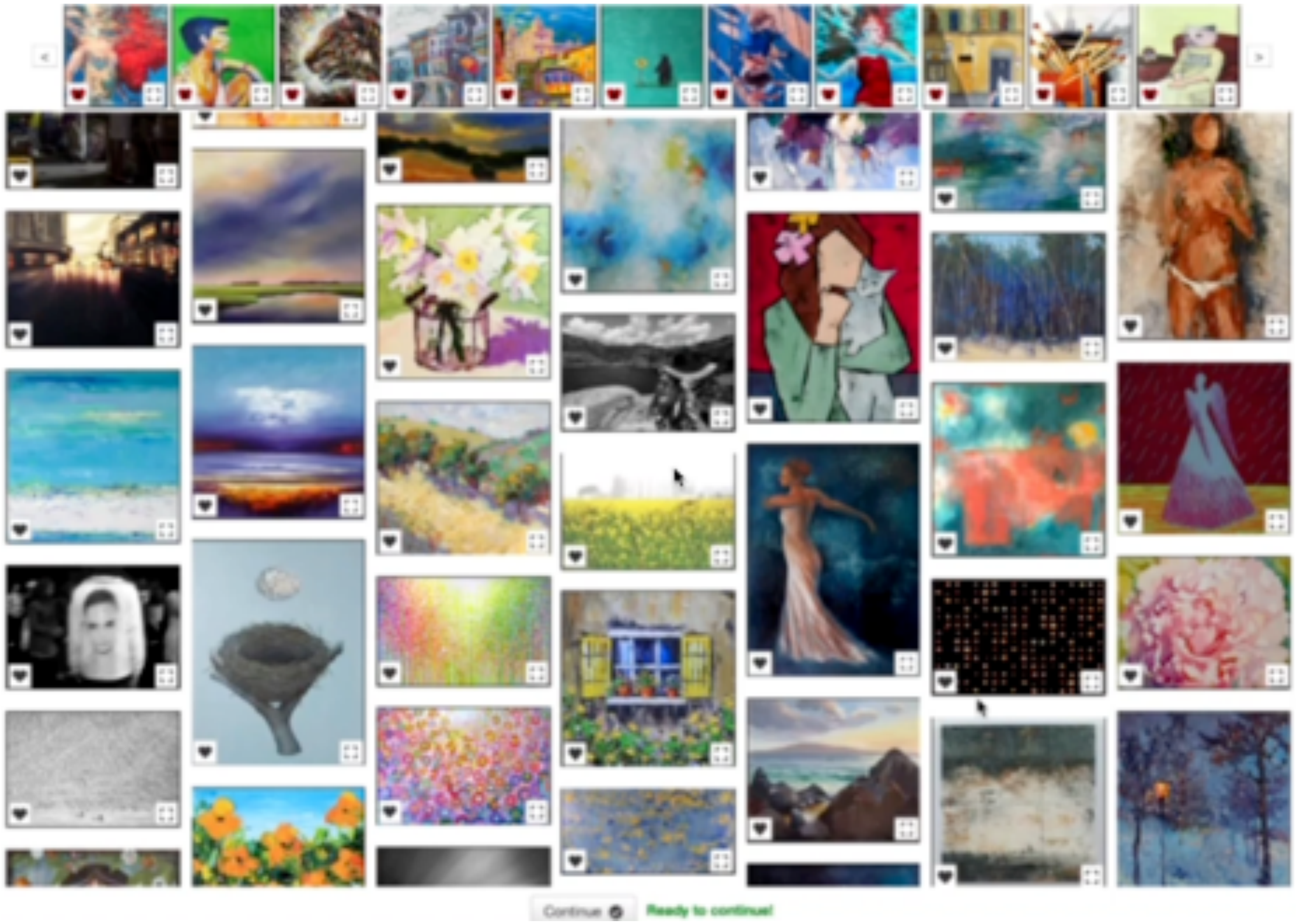
- CuratorNet improves upon VBPR and VisRank for the case of one-of-a-kind recommendation

- Unlike VBPR, CuratorNet does not need re-training to recommend to new users since it does not explicitly train user factors

- The proposed sampling guidelines benefit both CuratorNet and VBPR

Code and Experimental Data https://github.com/ialab-puc/CuratorNet

CuratorNet Demo Ugallery



Topic 2:

Brief intro to Deep Generative Models
+
Projects developed at PUC Chile

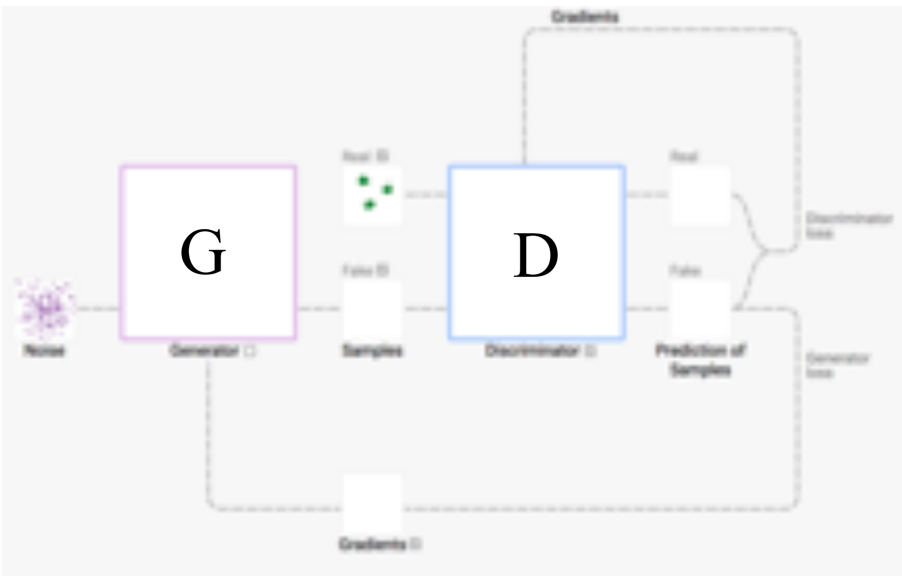
Generative Models in AI

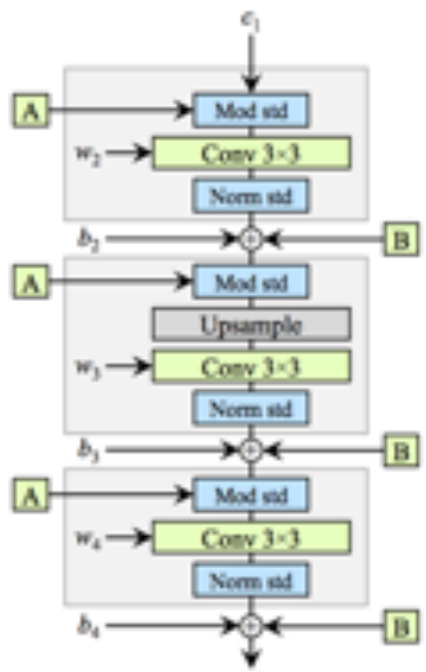
- Unlike discriminative models which are capable of classifying or predicting a number (regression), generative models are able to generate new instances not observed in the training data.

- In the latest years, models based on architectures like GANs and VAEs have yielded amazing results.

Example: face synthesis/generation

- There are several architectures, but the traditional GAN, once trained, can generate high-quality samples not observed in the training set.





(c) Revised architecture

Analyzing and Improving the Image Quality of StyleGAN https://arxiv.org/pdf/1912.04958.pdf

Topic 2:

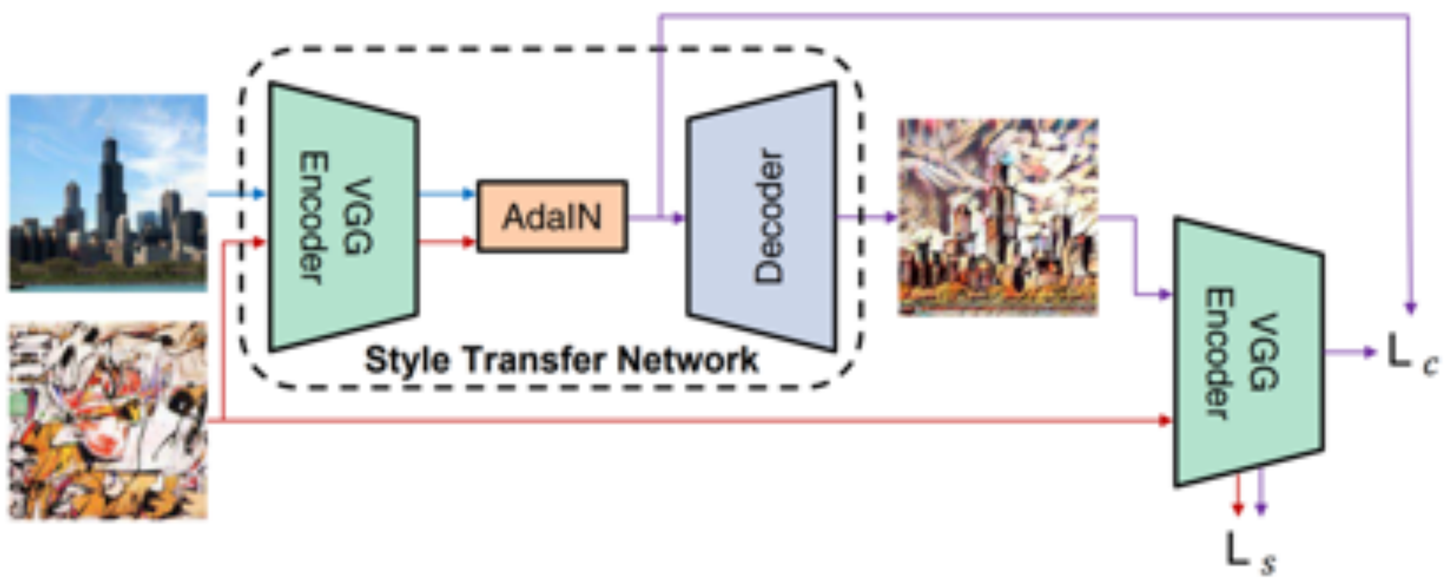
Brief intro to Deep Generative Models

+

Projects developed at PUC Chile

Style Transfer

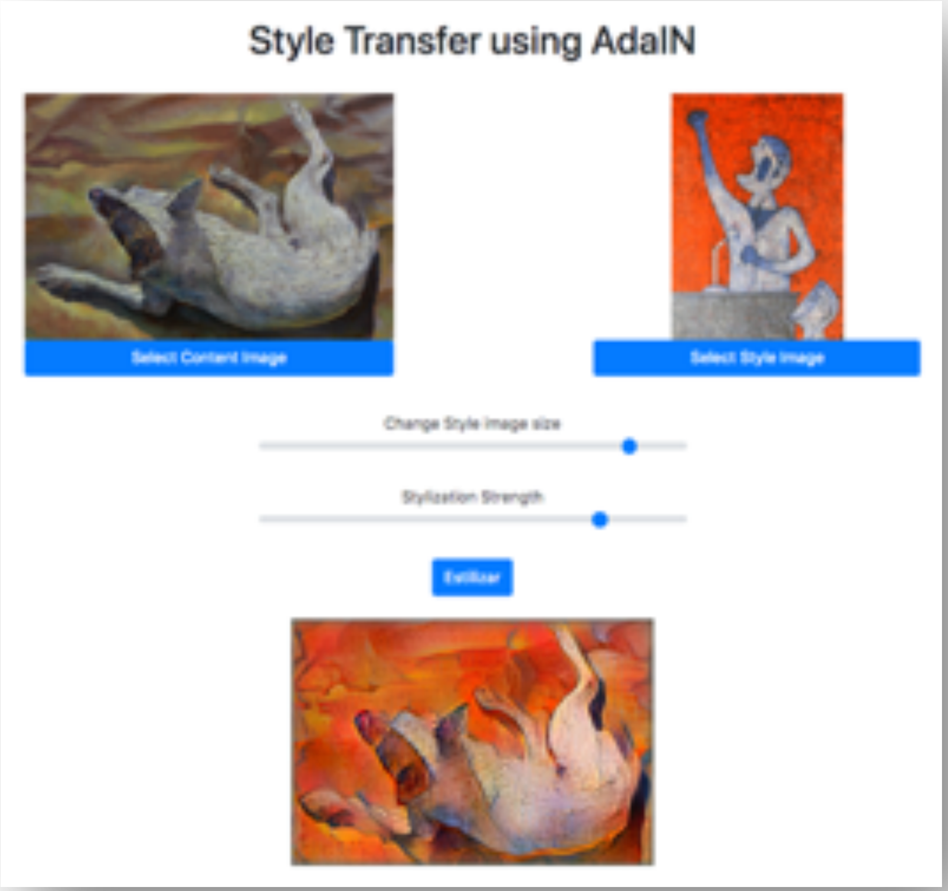
- Implementation from scratch of AdaIN:
  - Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1501-1510).



Jorge Perez Facuse

AdaIN UC

- https://niebla.ing.puc.cl/adain/



Ongoing Project

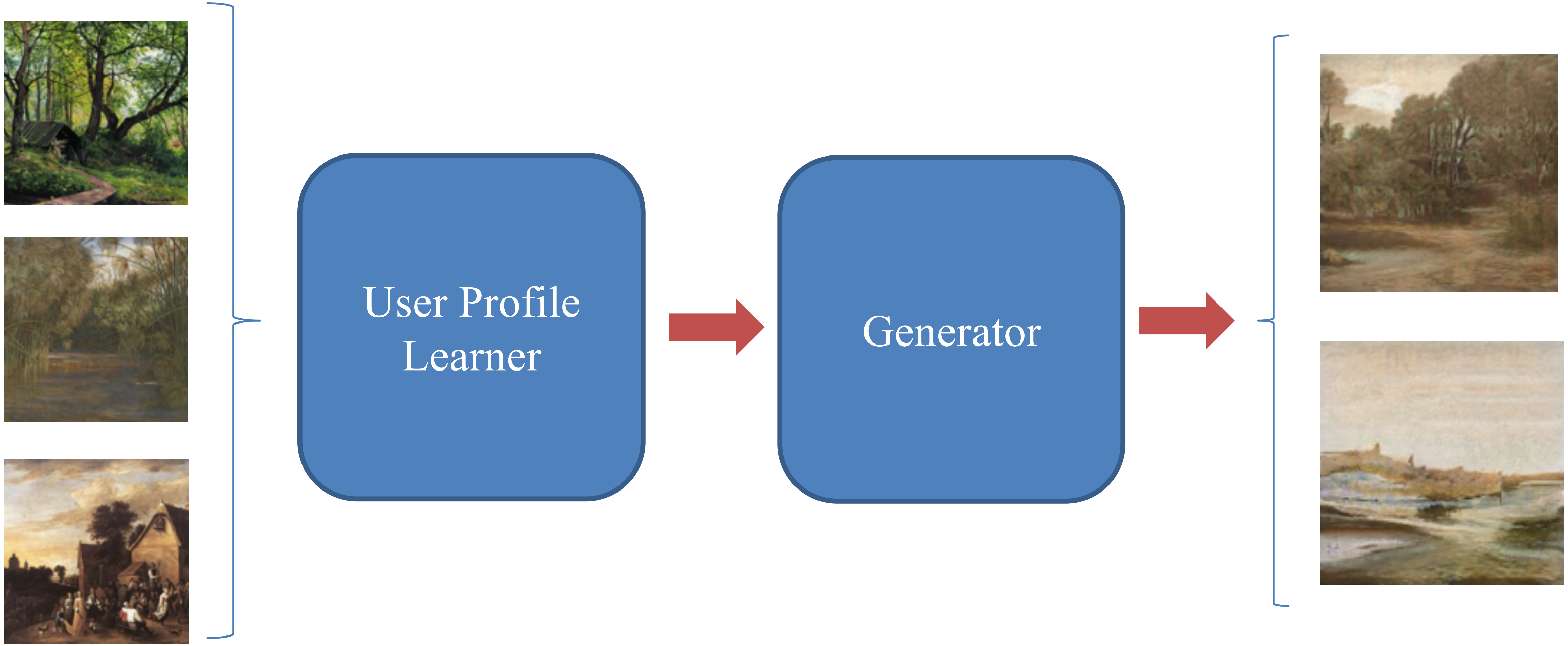
- Personalized generation of visual art: generative network conditioned on user preferences

Eugenio Herrera

step 0 step 147 step 521 step 613 step 804

Ongoing project

- Personalized recommendation by generation (real example!)



CreativAI UC Lab

https://creativai-uc.github.io/
https://github.com/CreativAI-UC/

CreativAI UC

- ArTeCiH project funded by PUC
- Co-directed by profesor Rodrigo Cádiz
- Purpose: Develop AI tools for music composition

- TimbreNet/Latent chords: recording of piano chords and implementation of a VAE for reconstruction and generation
  - We are studying the chords generated

Model: VAE with MFCC input

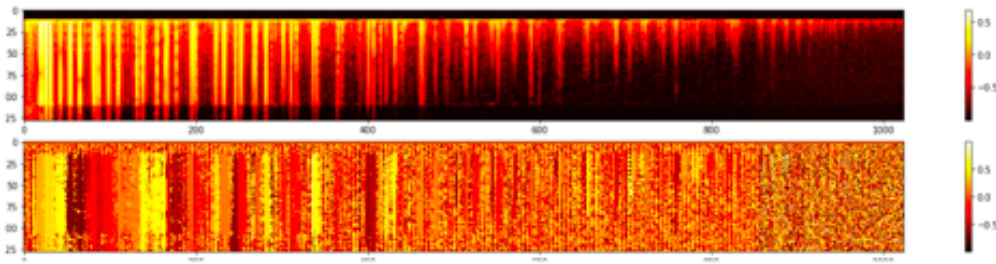


Figure 2: MFCC representation of a *forte* chord

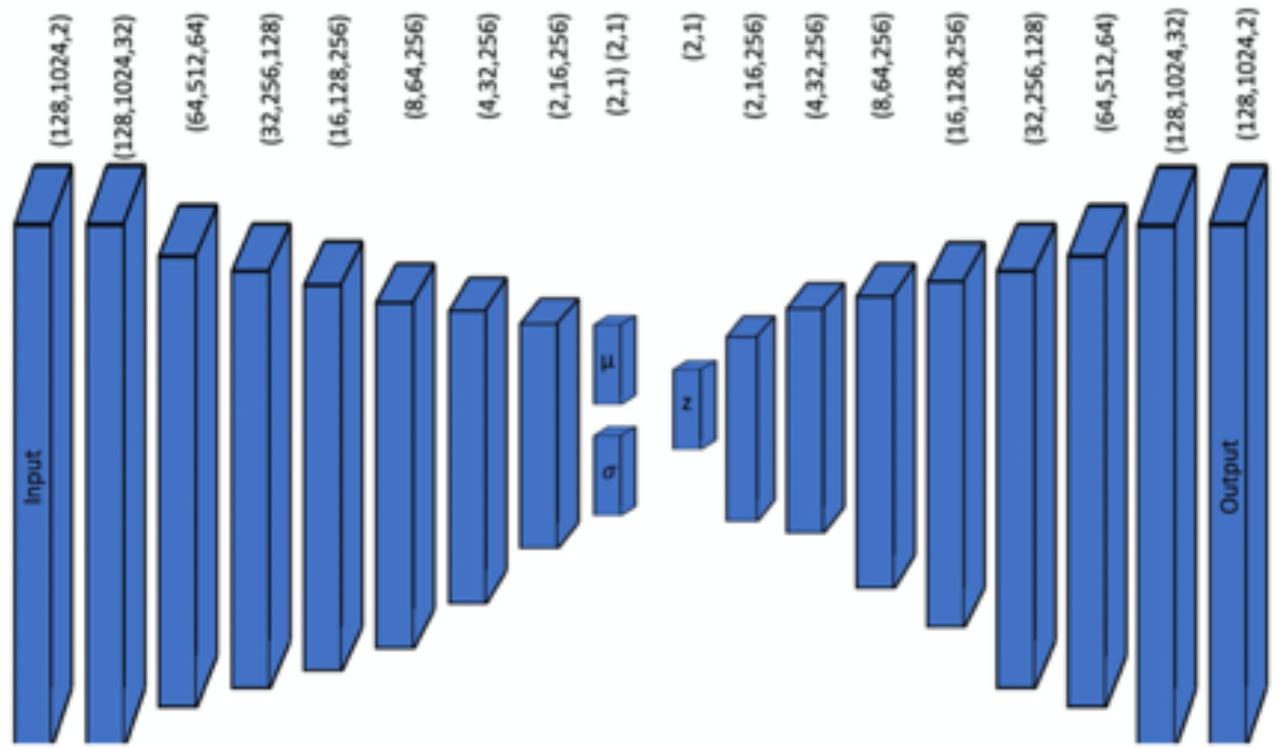


Figure 1: Arquitecture of our VAE model for chord synthesis.

Model: VAE with MFCC input

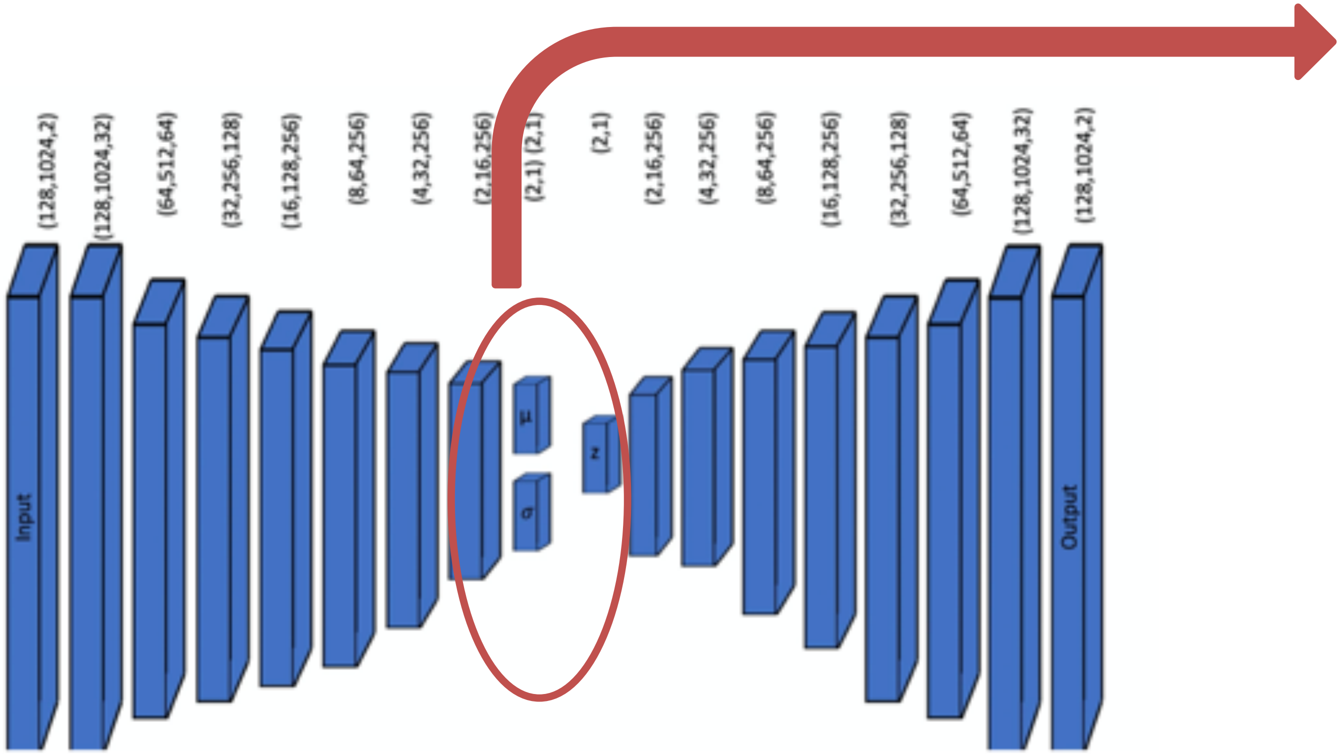


Figure 1: Arquitecture of our VAE model for chord synthesis.

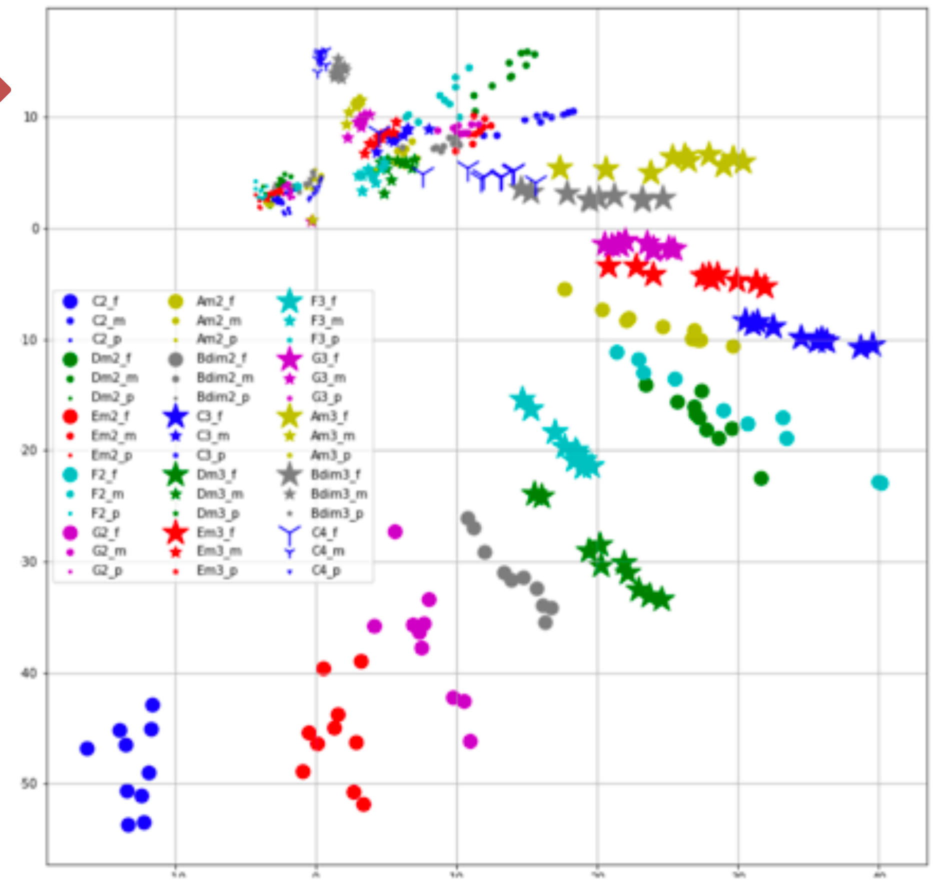
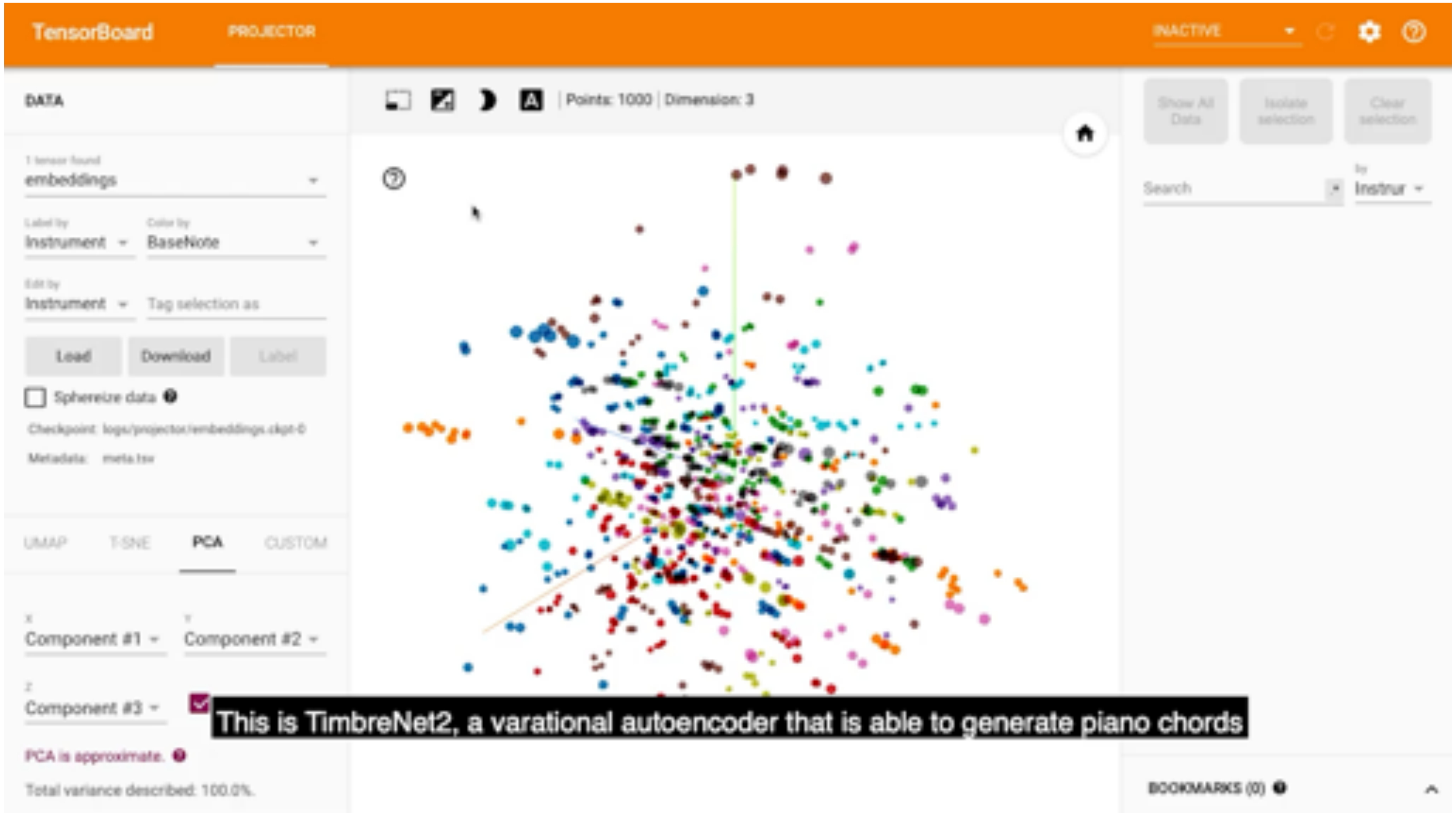


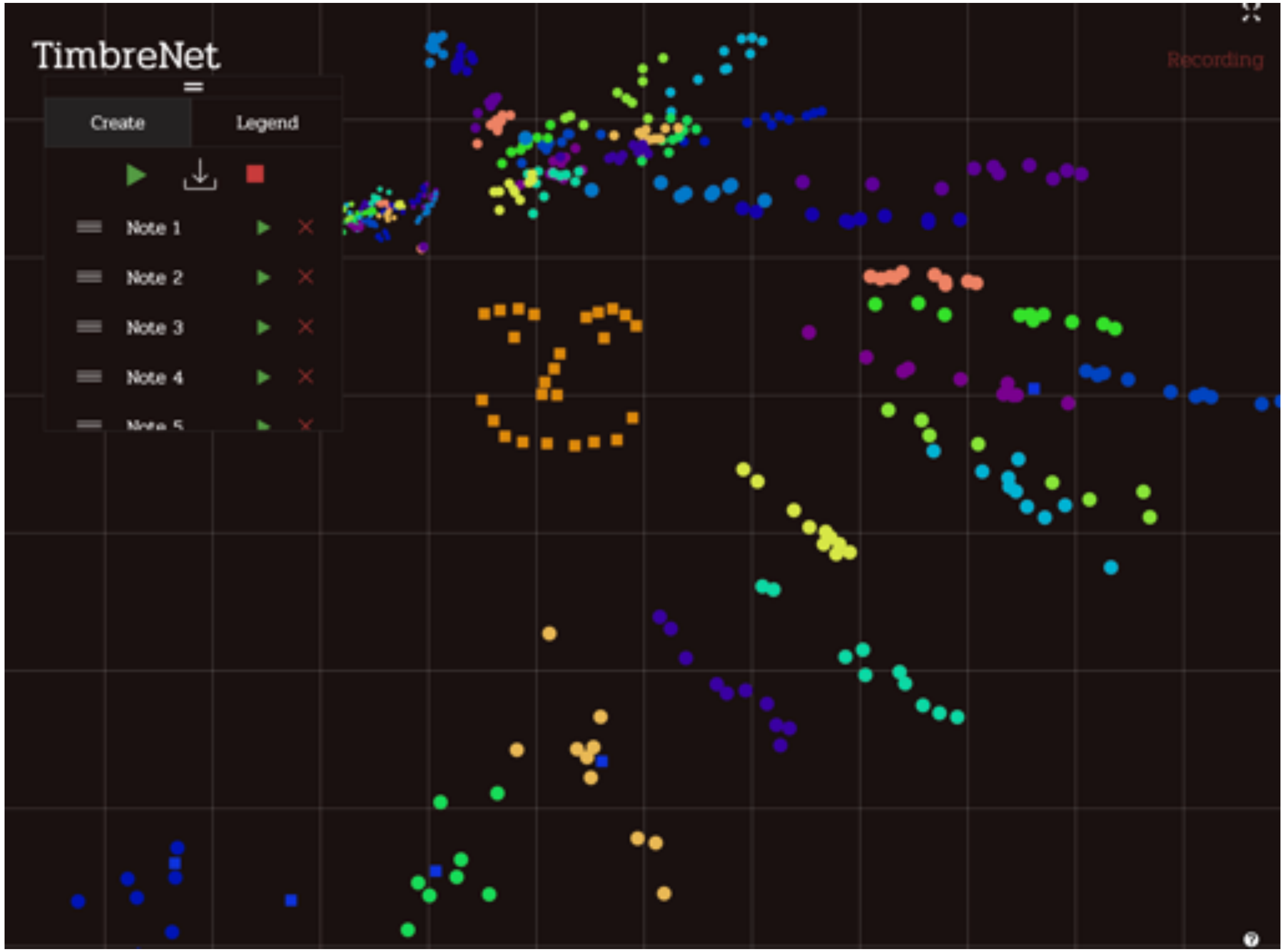
Figure 4: Two dimensional latent space representation of the dataset. Chords are arranged in a spiral pattern, and chords are arranged from forte to a piano dynamic.

TimbreNet Demo

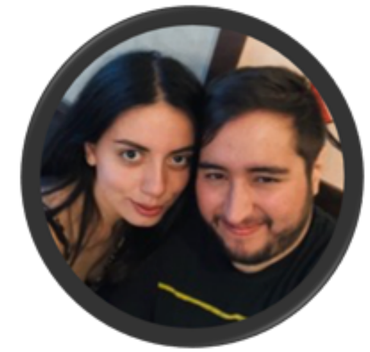


Agustín Macaya

http://timbreplay.ml/

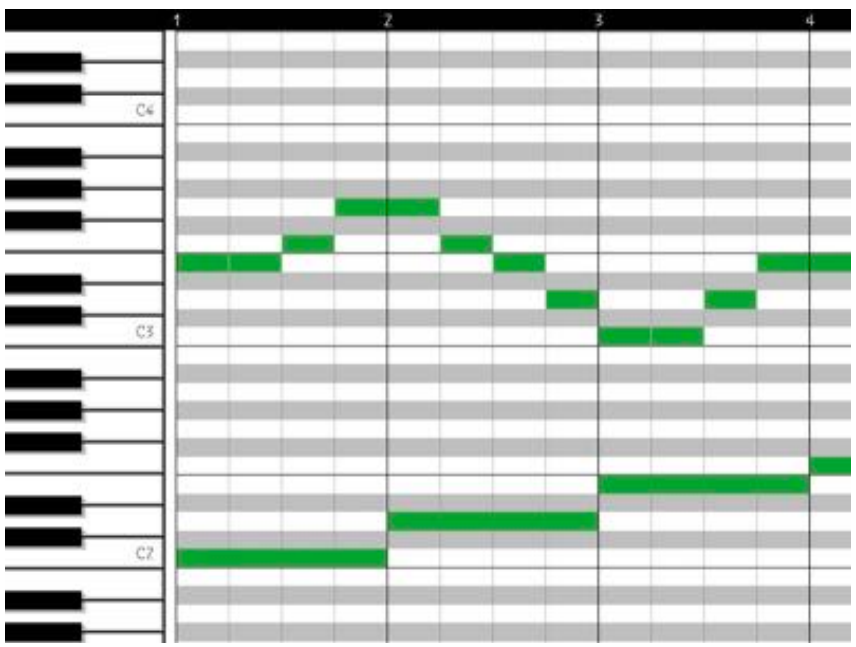


GAN Piano Rolls

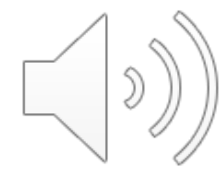


Manuel Cartagena

- MIDI Generation from Piano rolls
- Few samples after 10 and 300 epochs



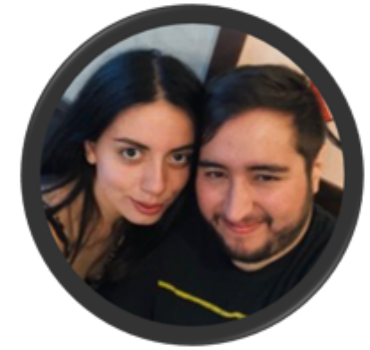
https://w00zie.github.io/post/wae/



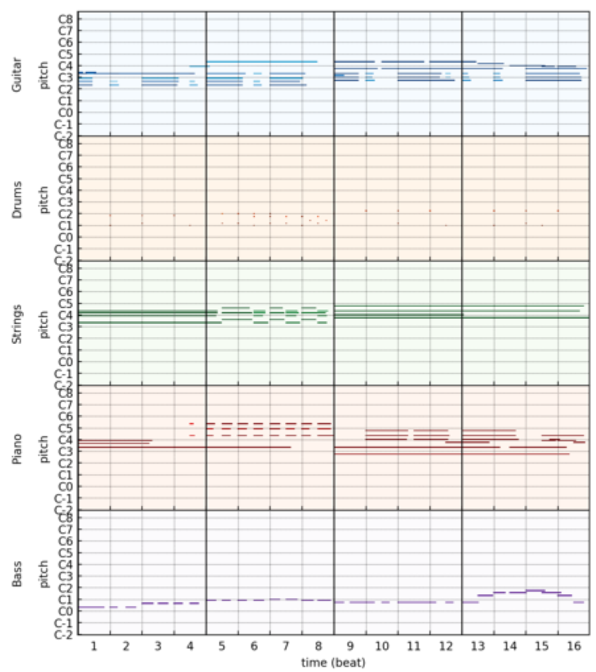
GAN Piano Rolls

- MIDI Generation from Piano rolls

- Now: working on multitrack generation!



Manuel Cartagena

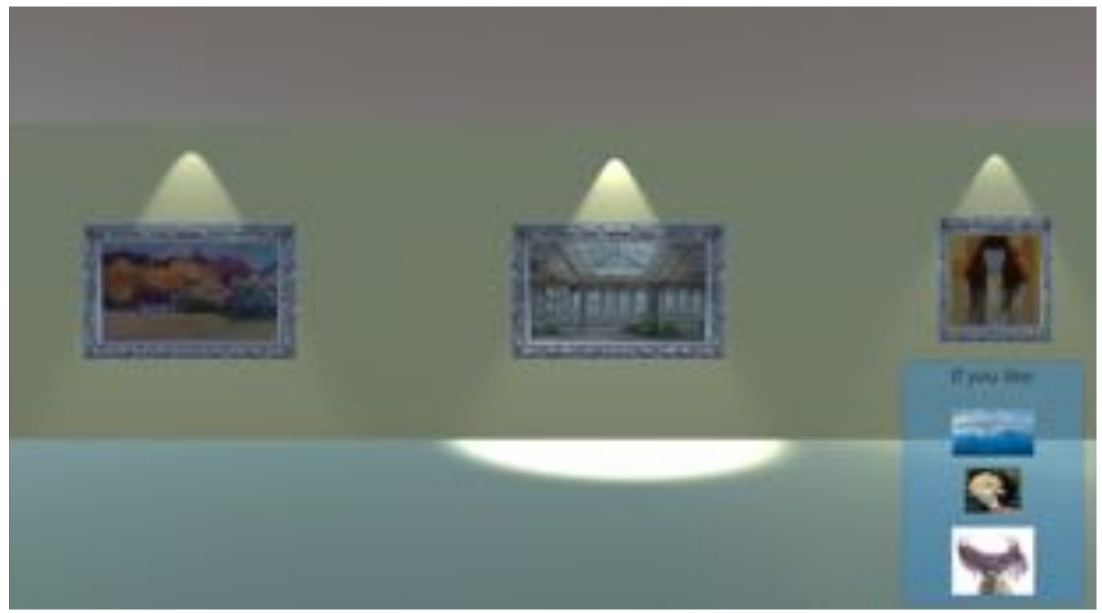


https://w00zie.github.io/post/wae/

Other projects

- Visual art recommendation on immersive environments

Isidora Palma



Summary and discussion

- Tools to amplify possibilities of artistic creativity.

- Attribution on generative models: should creators of the samples used to create these models be also given credit ?

- Fair recommendation: not only for users/consumers, also for content creators.

Can computers create art ?

- https://www.mdpi.com/2076-0752/7/2/18

Can Computers Create Art?†

by Aaron Hertzmann

Adobe Research, San Francisco, CA 94103, USA

† This essay expresses my own opinions and not those of my employer.

Arts **2018**, *7*(2), 18; https://doi.org/10.3390/arts7020018

Received: 4 February 2018 / Revised: 5 May 2018 / Accepted: 7 May 2018 / Published: 10 May 2018

(This article belongs to the Special Issue The Machine as Artist (for the 21st Century))

View Full-Text | Download PDF | Browse Figures | Cite This Paper

Abstract

This essay discusses whether computers, using Artificial Intelligence (AI), could create art. First, the history of technologies that automated aspects of art is surveyed, including photography and animation. In each case, there were initial fears and denial of the technology, followed by a blossoming of new creative and professional opportunities for artists. The current hype and reality of Artificial Intelligence (AI) tools for art making is then discussed, together with predictions about how AI tools will be used. It is then speculated about whether it could ever happen that AI systems could be credited with authorship of artwork. It is theorized that art is something created by social agents, and so computers cannot be credited with authorship of art in our current understanding. A few ways that this could change are also hypothesized. View Full-Text

Keywords: art and technology; artificial intelligence; photography; painting; computer animation; image stylization

▼ Show Figures

Thanks!

- dparra@ing.puc.cl
- https://creativai-uc.github.io/
- http://dparra.sitios.ing.uc.cl/

The effect of explanations on Art RecSys

- DNN features: good performance, but difficult to make feature explanations
- Manually-engineered features: not strong performance, but easier to explain
- What has stronger effect upon user trust and satisfaction?

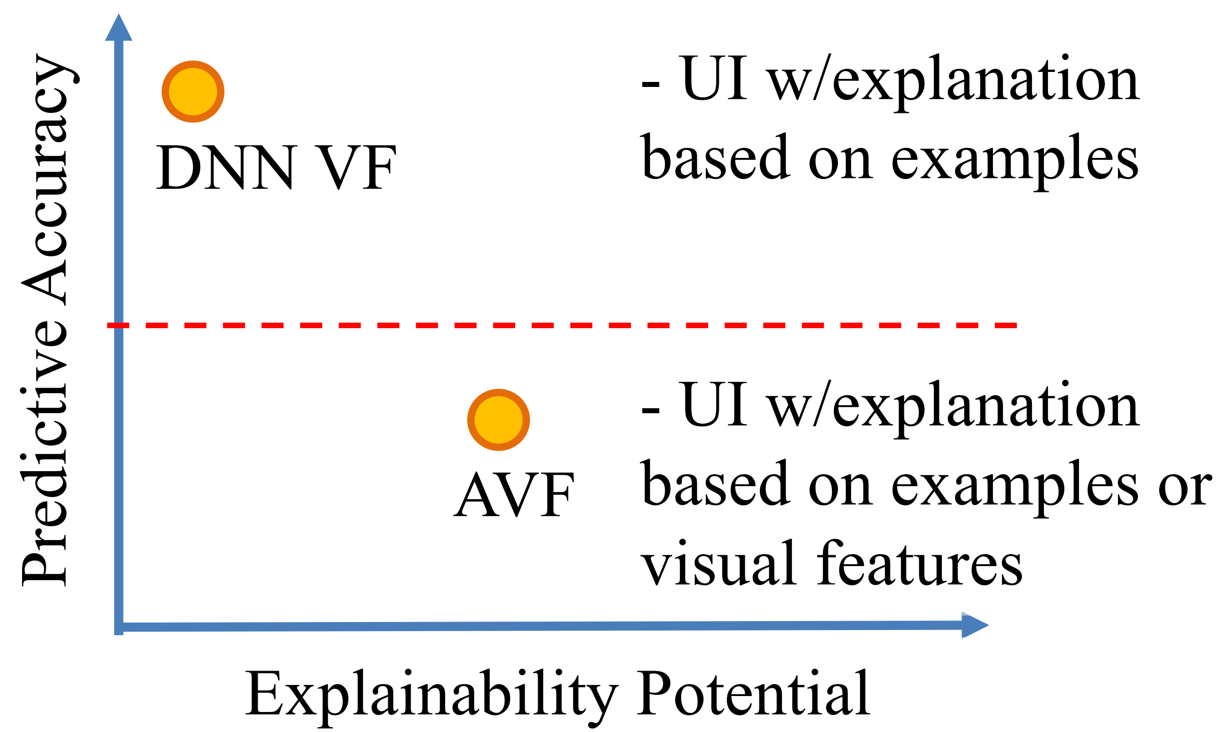
XAI for visual art RecSys

Dominguez, V., Messina, P., Donoso-Guzmán, I., & Parra, D. (2019). **The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images**. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 408-416).

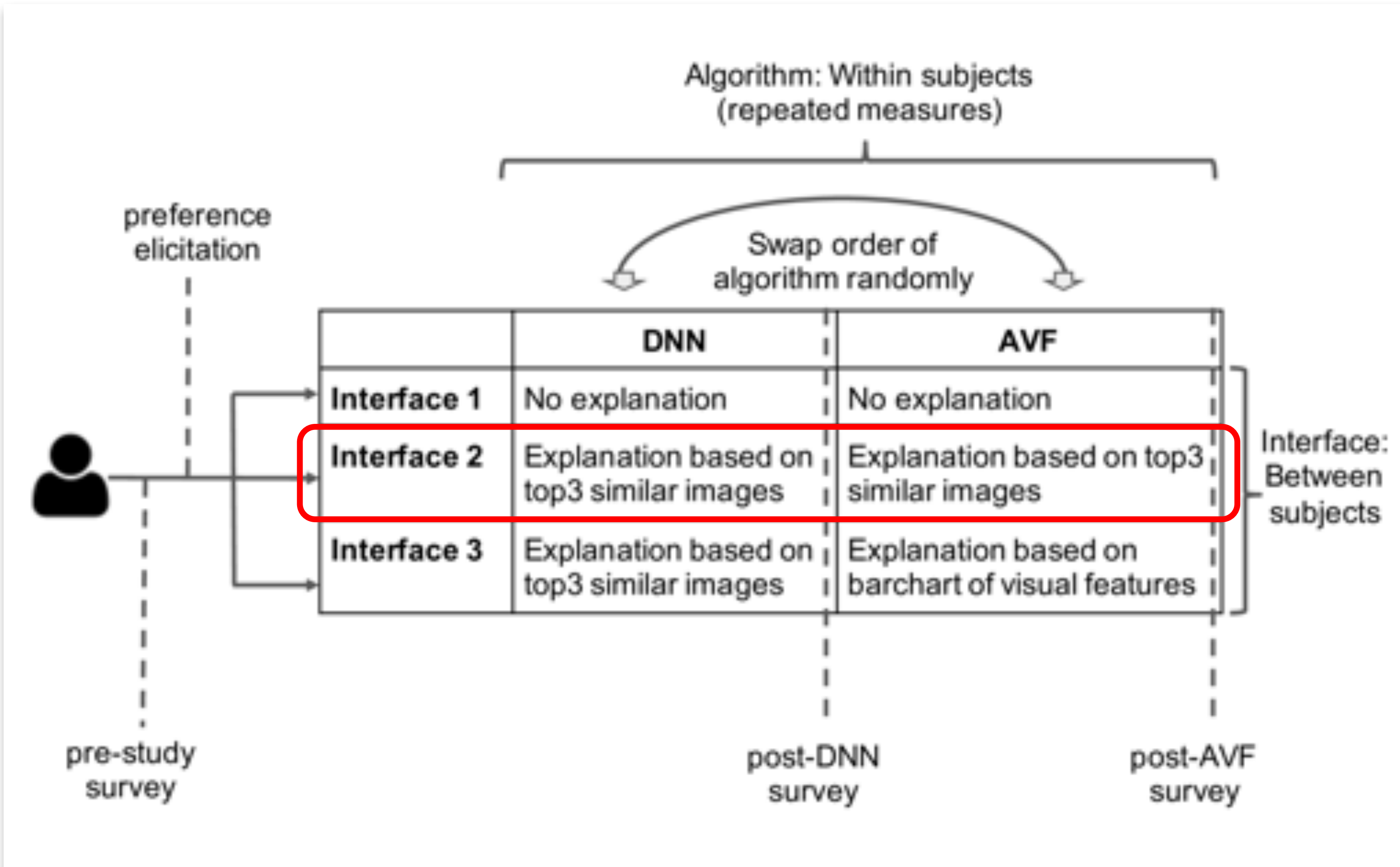
* Outstanding paper award

Our General Research Question

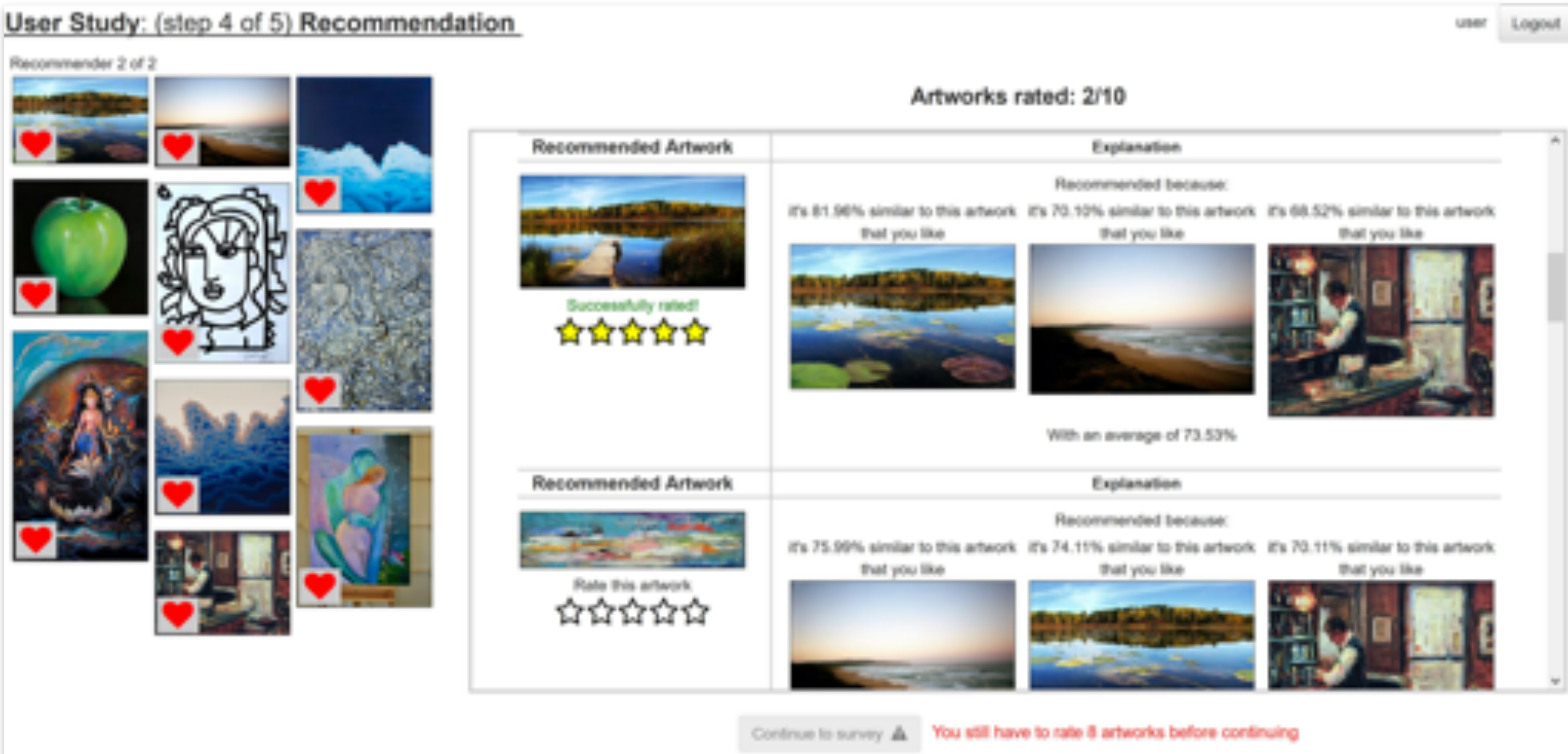
- Explainable & Transparent UIs are well perceived by users in previous RecSys research, but: Does it matter to UX if you do not produce accurate enough predictions?



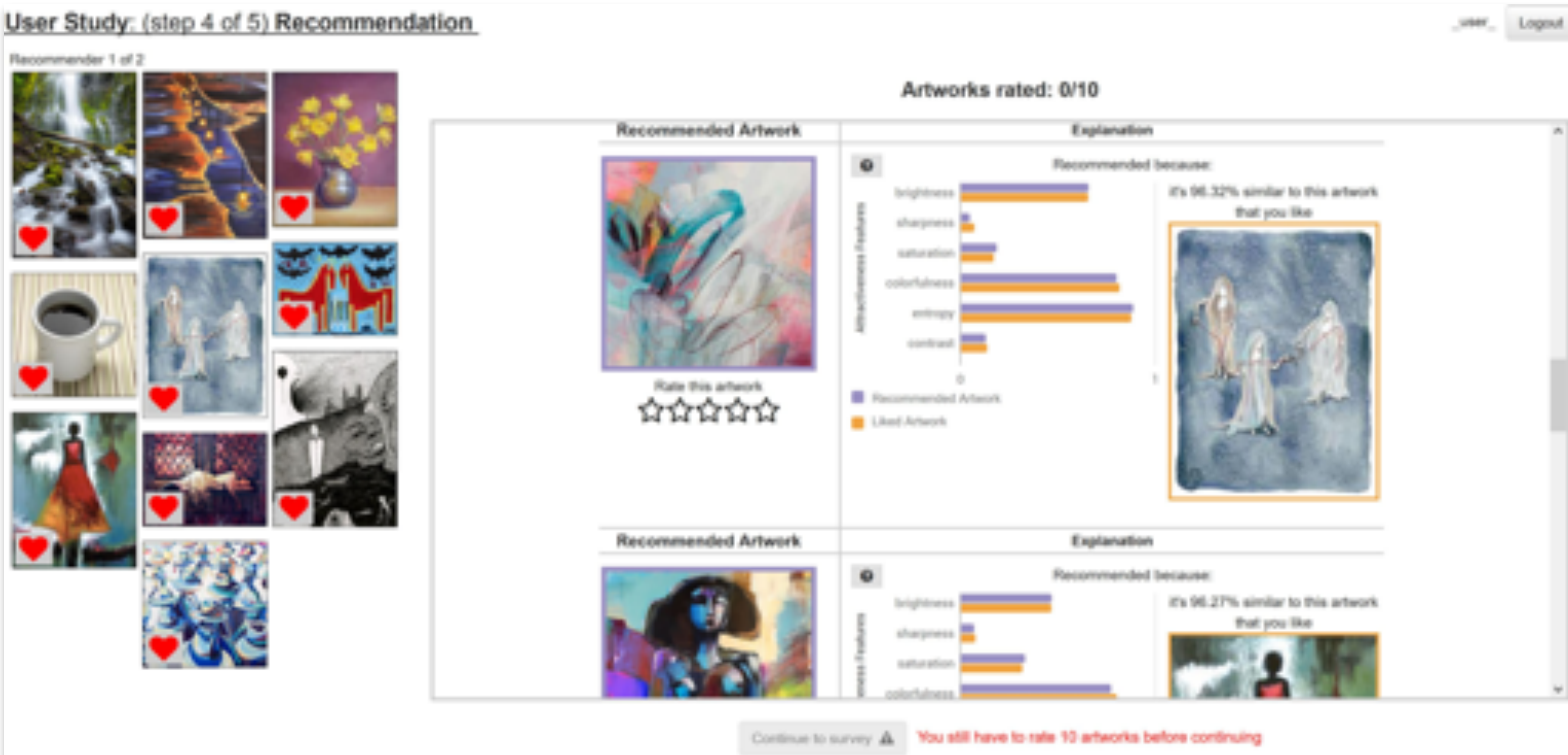
Study Procedure



Interface 2: explainable, no transparency



Interface 3: explainable & transparent



Evaluation & Results

Study on Amazon Mechanical Turk:

- 121 valid users completed correctly the study.
- Task took them around 10 minutes to complete.
- ~56% female, 44% male.
- 80% attended to 1 or more art classes at high school level or above.
- 80% visited museums or art galleries at least once a year.

Results

	Evaluation Dimensions													
	Explainable		Relevance		Diverse		Interface Satisfaction		Use Again		Trust		Average Rating	
Condition	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF
Interface 1 (No Explanations)	66.2*	51.4	69.0*	53.6	46.1	69.4*	69.9	62.1	65.8	59.7	69.3	63.7	3.55*	3.23
Interface 2 (DNN & AVF: Top-3 similar images)	83.5*†[1]	74.0†[1]	80.0*	61.7	58.8	69.9*	76.6*	61.7	76.1*	65.9	75.9*	62.7	3.67*	3.00
Interface 3 (DNN: Top-3 similar, AVF: feature bar chart)	84.2*†[1]	70.4†[1]	82.3*†[1]	56.2	65.3†[1]	71.2	69.9*	63.3	78.2*	58.7	77.7*	55.4	3.90*	2.99

Interface 1: UI without explanation
Interface 2: UI with example-based explanation
Interface 3: UI with transparent explanation (AVF)

Results

	Evaluation Dimensions													
	Explainable		Relevance		Diverse		Interface Satisfaction		Use Again		Trust		Average Rating	
Condition	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF	DNN	AVF
Interface 1 (No Explanations)	66.2*	51.4	69.0*	53.6	46.1	69.4*	69.9	62.1	65.8	59.7	69.3	63.7	3.55*	3.23
Interface 2 (DNN & AVF: Top-3 similar images)	83.5*†[1]	74.0†[1]	80.0*	61.7	58.8	69.9*	76.6*	61.7	76.1*	65.9	75.9*	62.7	3.67*	3.00
Interface 3 (DNN: Top-3 similar, AVF: feature bar chart)	84.2*†[1]	70.4†[1]	82.3*†[1]	56.2	65.3†[1]	71.2	69.9*	63.3	78.2*	58.7	77.7*	55.4	3.90*	2.99

7 dimensions evaluated, for DNN and AVF (scale 1-100):

Perception of:
- Explainability
- Relevance
- Diversity
- Satisfaction w/UI
- Intention of use
- Trust on RecSys
- Avg. Rating