

# AI-cyclopedia: A look into fascinating art and understanding why they are what they are

Shaoyang Luo

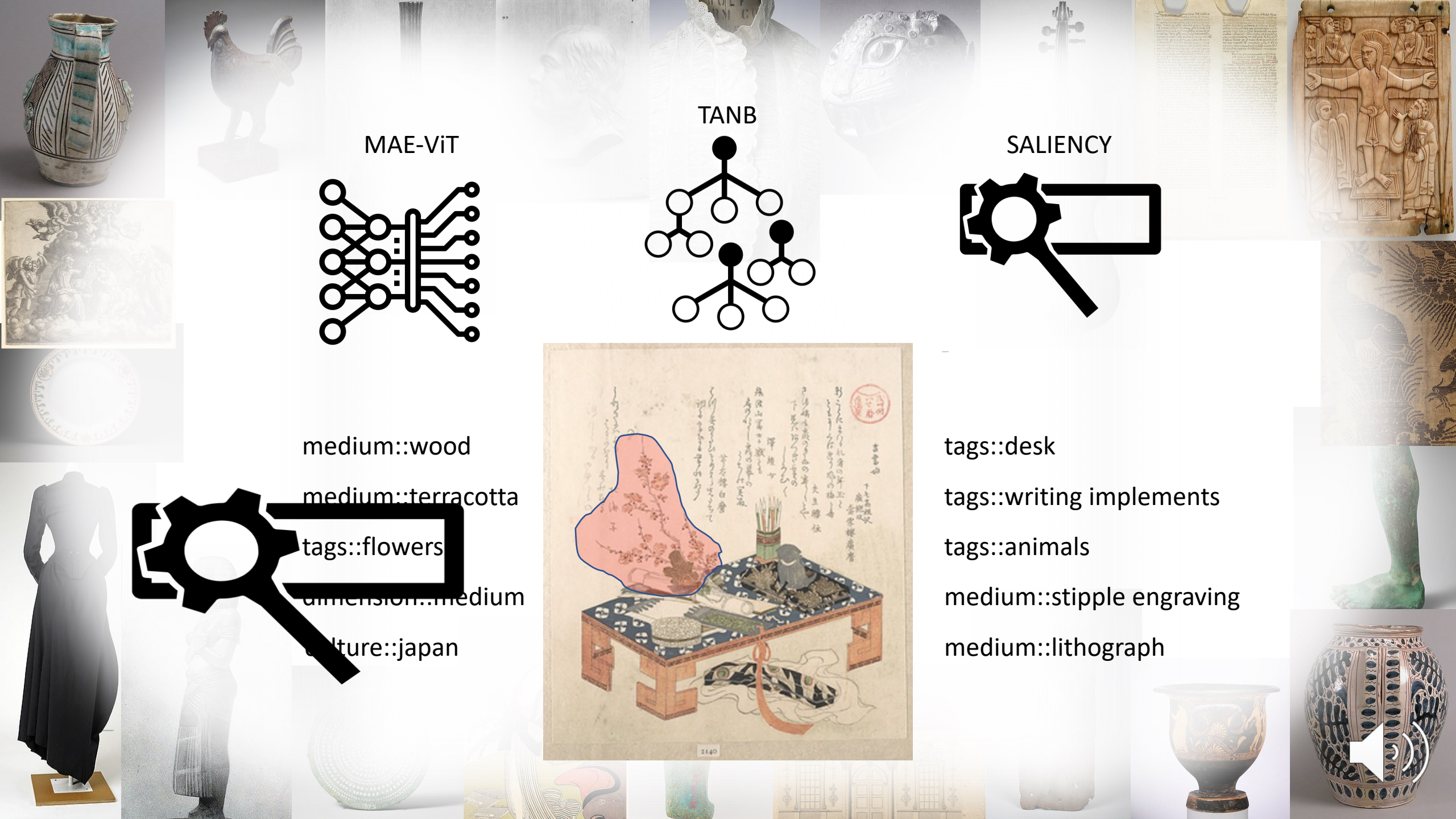
Roman Puttkammer

Kana Sethu

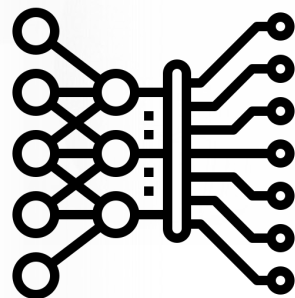
{sluo3, rputtkam, ksethu}@Stanford.edu



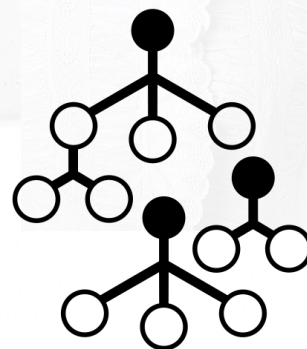




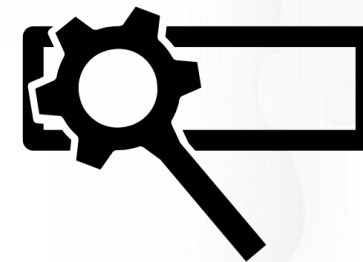
MAE-ViT



TANB



SALIENCY



medium::wood

medium::terracotta

tags::flowers

dimension::medium

culture::japan



tags::desk

tags::writing implements

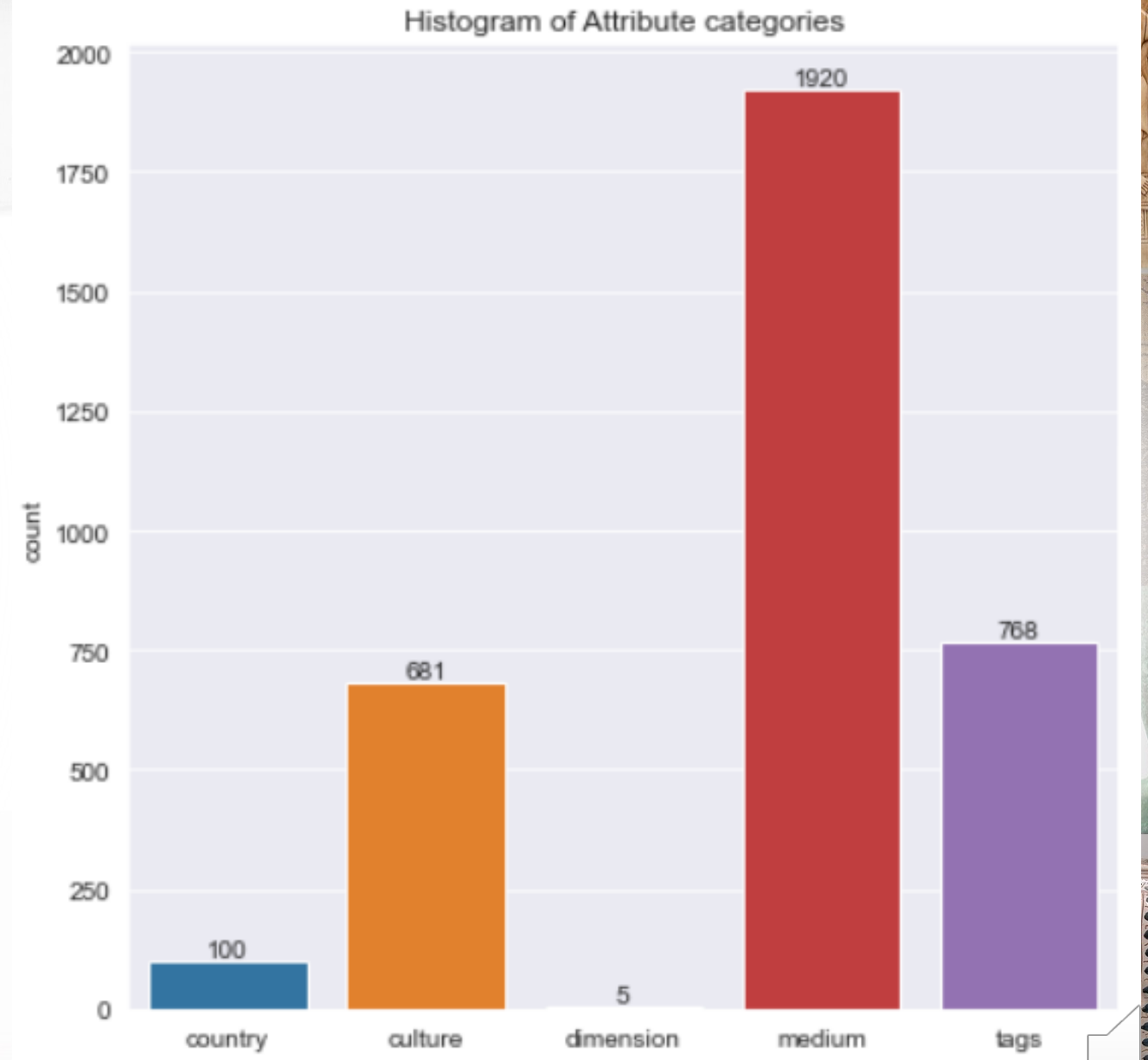
tags::animals

medium::stipple engraving

medium::lithograph

# IMET2021FGVC8 Dataset

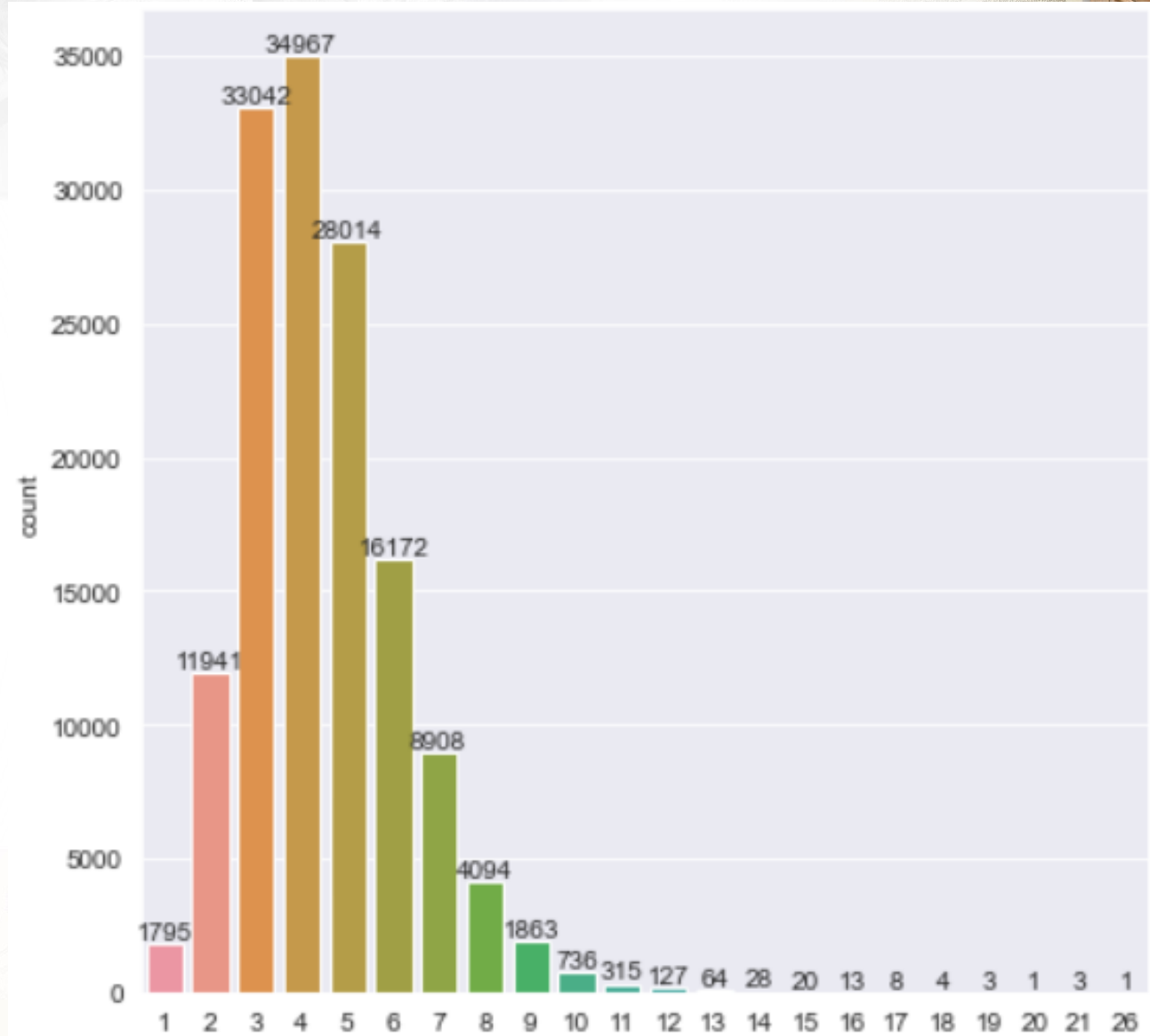
- Train: 145K images
- Test: 53K
- Unique Attributes: 3475





# Histogram of attributes

- Most images have 6
- But some can have as much as 20!
- Many of which can be conflicting



# Image Classification with DNN

## Experiments

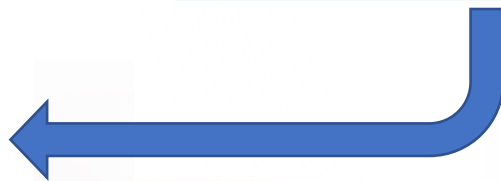
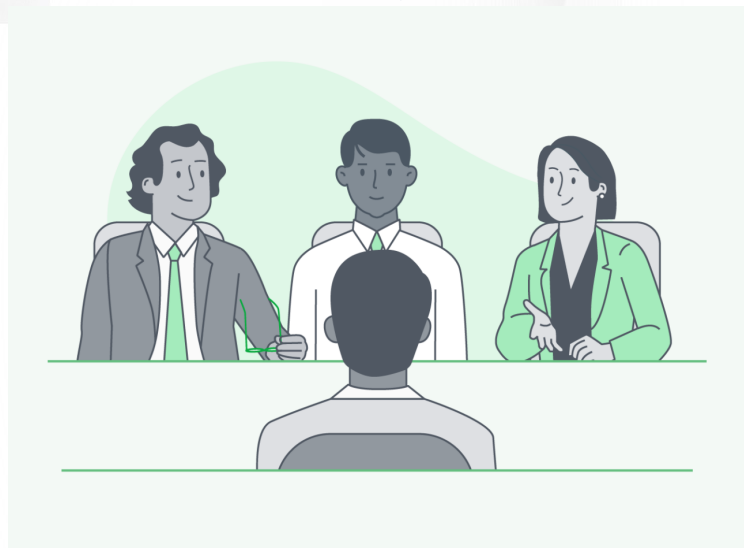
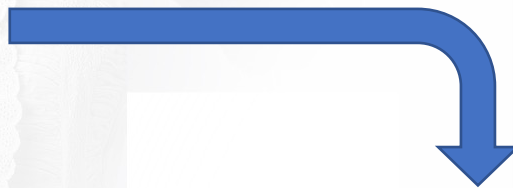
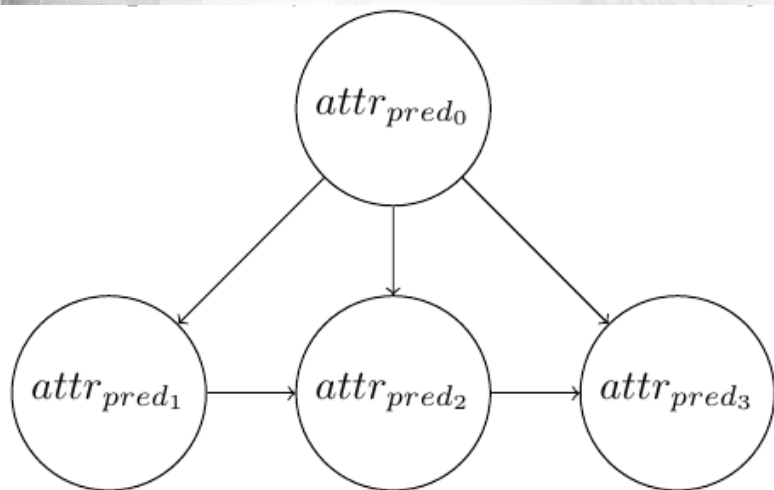
- 3 CNN Architectures
- Metrics: PR-AUC, F2
  - BCE LOSS

## Results

Model	Train	Validation	Test
	Weighted PR-AUC		
ViT	.95	.59	.59
Simple CNN	.40	.40	.39
EfficientNET	.41	.37	.37

Model	F2	F2
	Threshold=0.5	Threshold=OPT
ViT	.23	.28
Simple CNN	.14	.15
EfficientNET	.11	.16





TopN Attribute	F2 Score
4	0.74
5	0.72
6	0.70

hay



quail



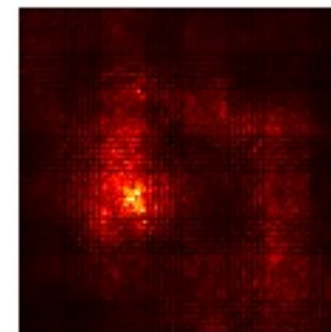
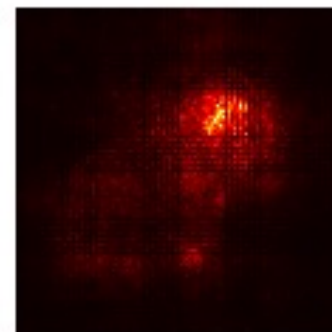
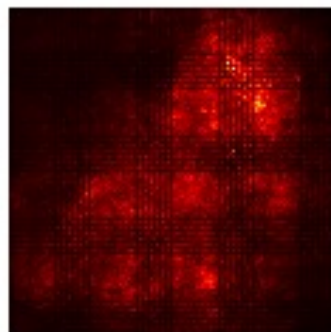
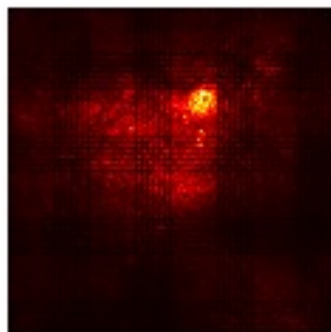
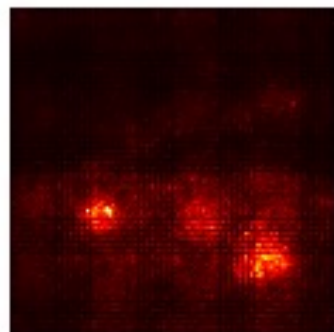
Tibetan mastiff



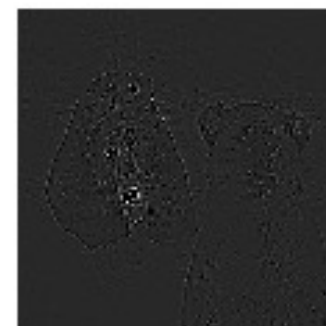
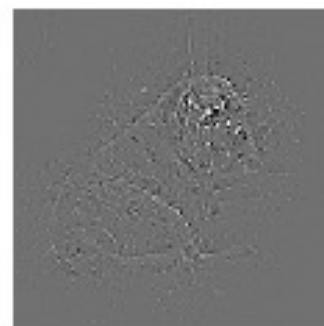
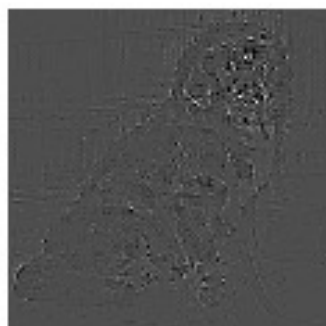
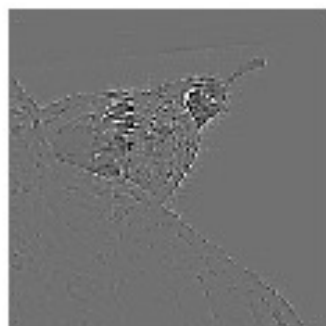
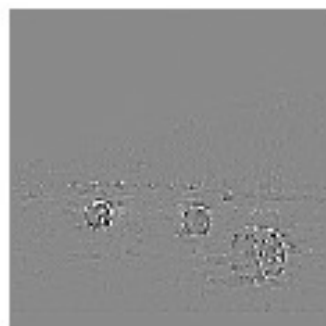
Border terrier



brown bear, bruin, Ursus arctos



(b) Smooth Grad



(d) Guided Backprop



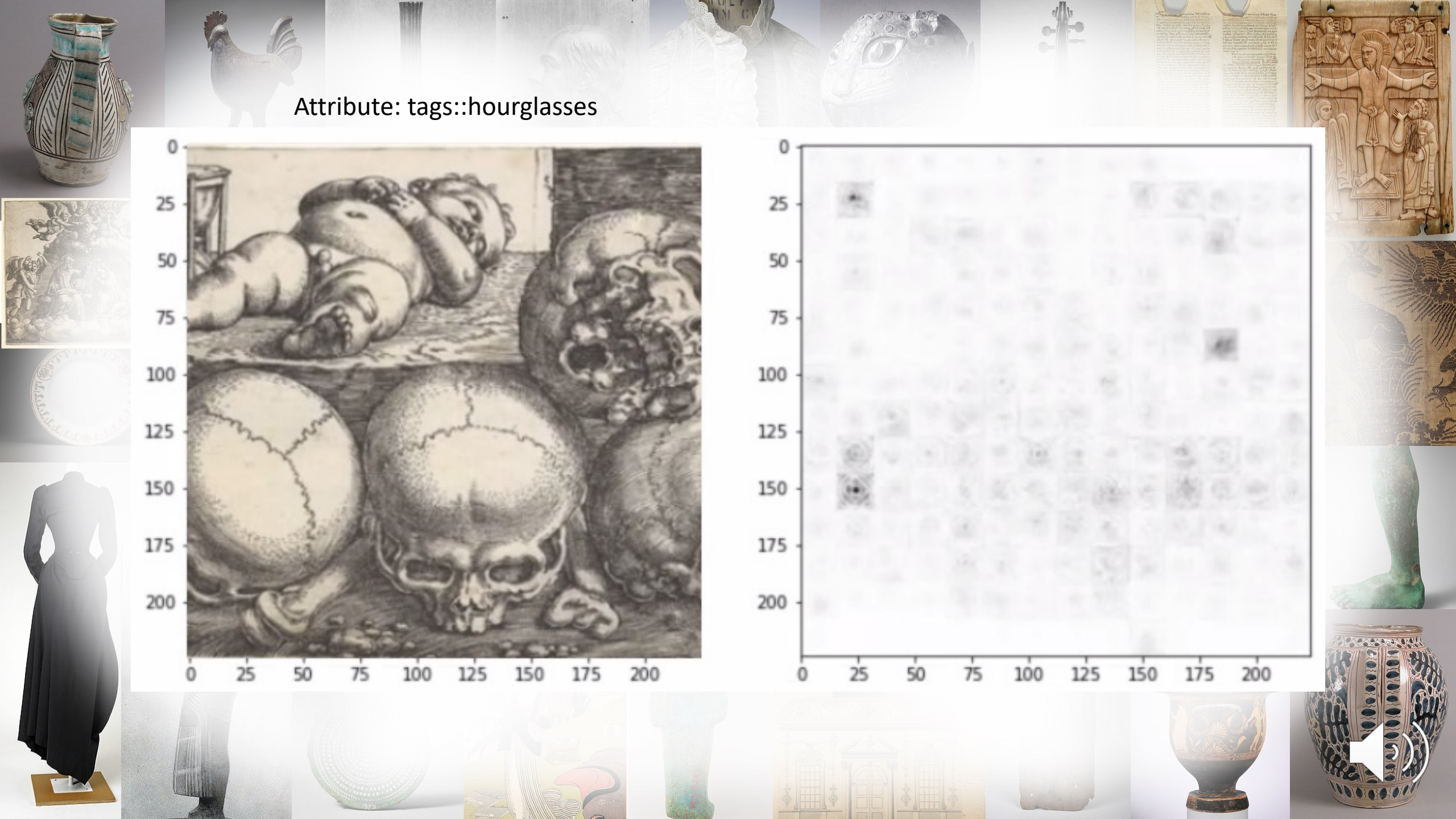




Attribute:  
medium::glaze



Attribute: tags::hourglasses



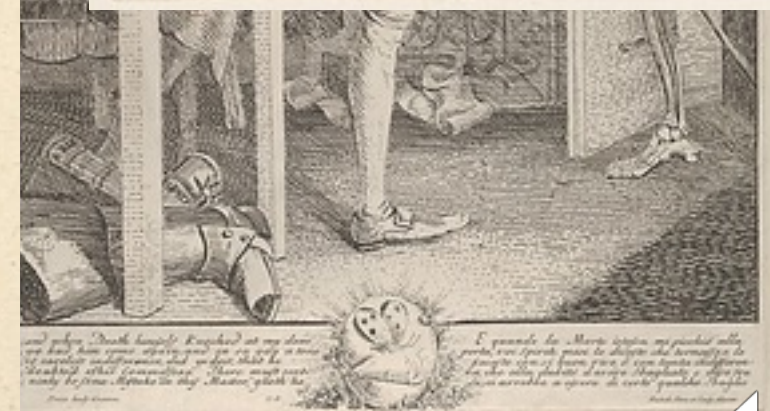




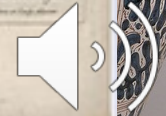
Homo natus de muliere, breui vivens tem-  
pore, repletur multis miserijs: qui quasi flos  
erigitur & conteritur & fugit velut vmbra



TEMPS EGO IMMENSVM SPATIES DIMETTOR ORBEM



E quando la Morte inghia ne pancha alla  
porta, non parca mai la disparte che dormiva la  
figlia, non si fuo per il suo letto che dormiva  
in che non parca di averla dormita e non fu  
che non parca di averla dormita e non fu  
che non parca di averla dormita e non fu

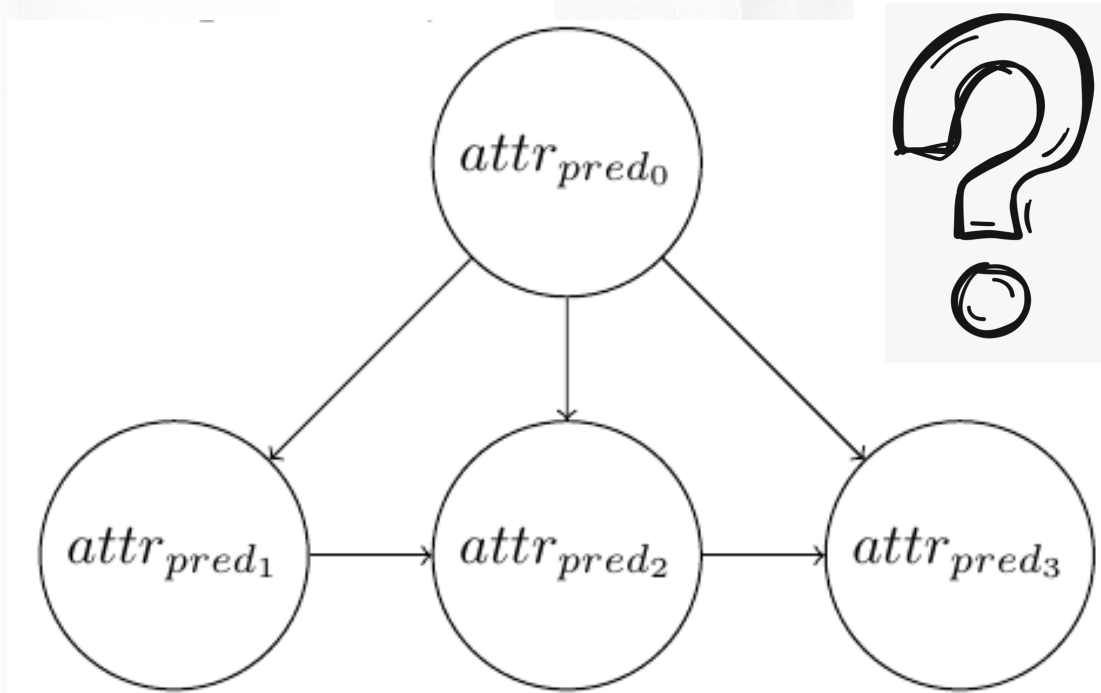
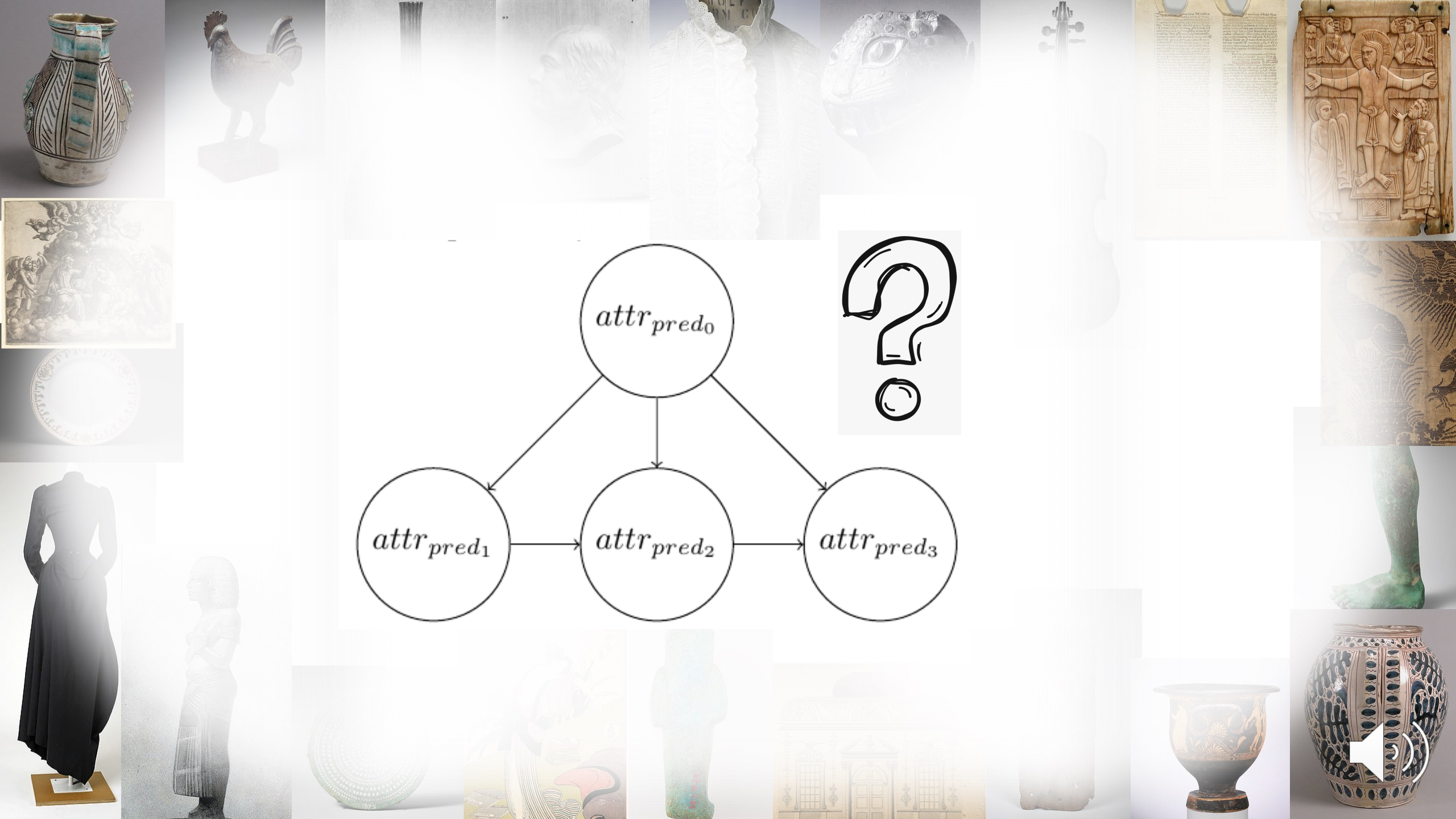




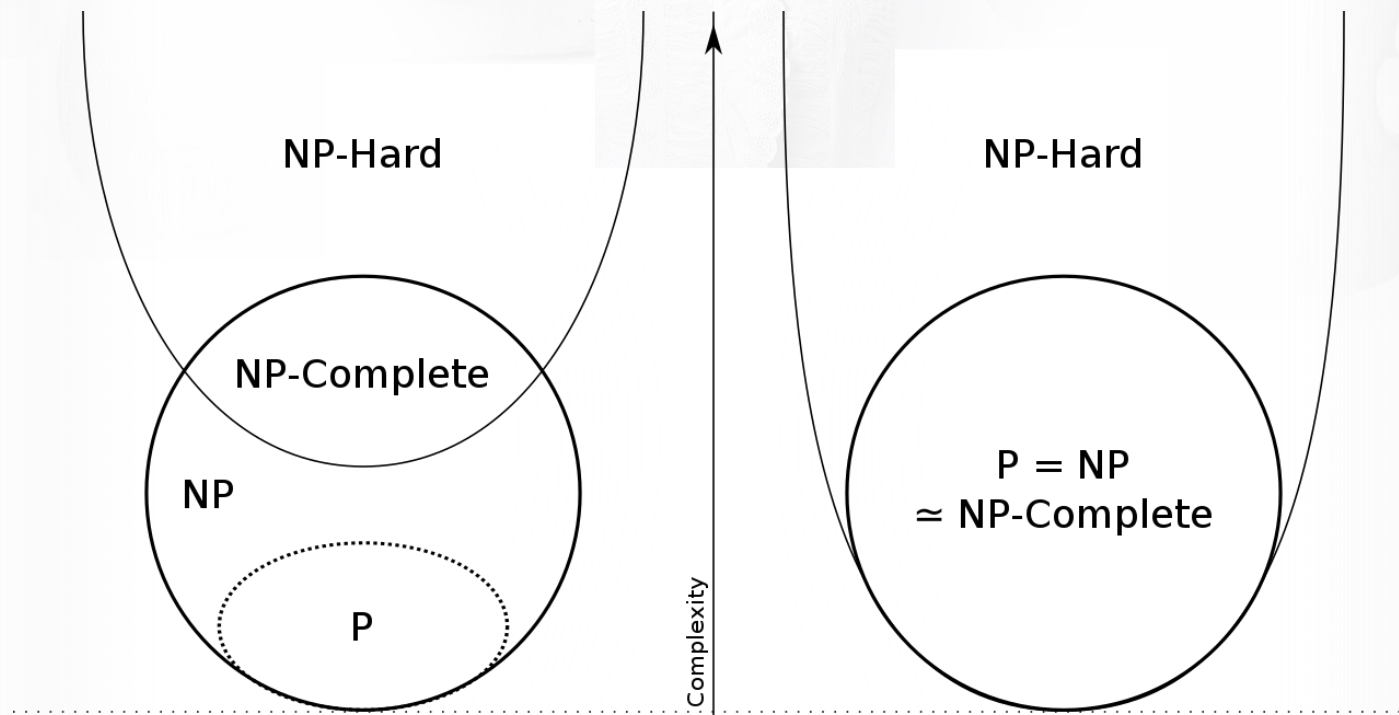
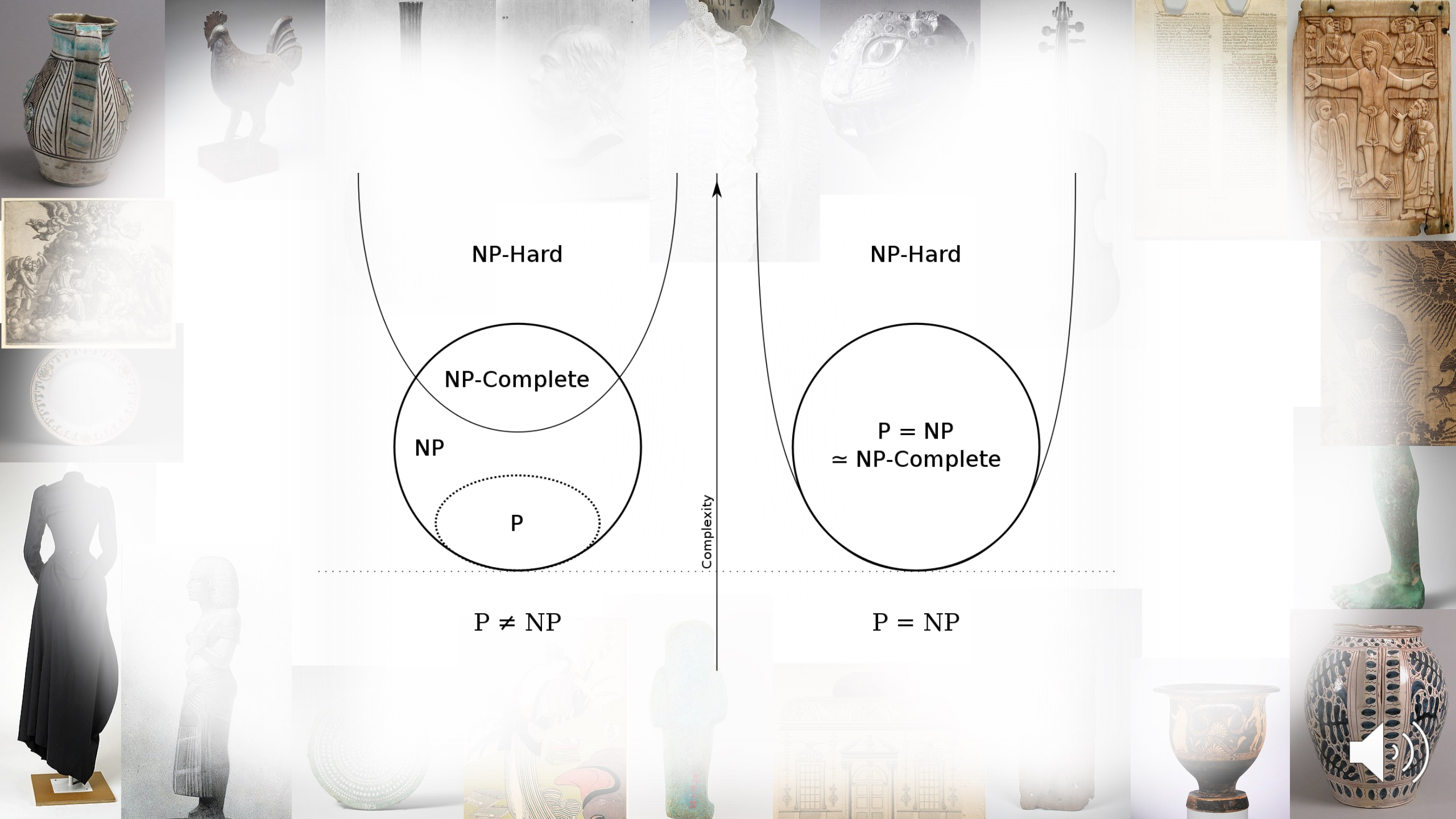
## Future Work

- Self-supervised pretraining of ViTMAE to improve latent representations.
- Alternate/improved data augmentation.
- ViT-specific methods for saliency.

(See project report for all references.)







NP-Hard

NP-Hard

NP-Complete

NP

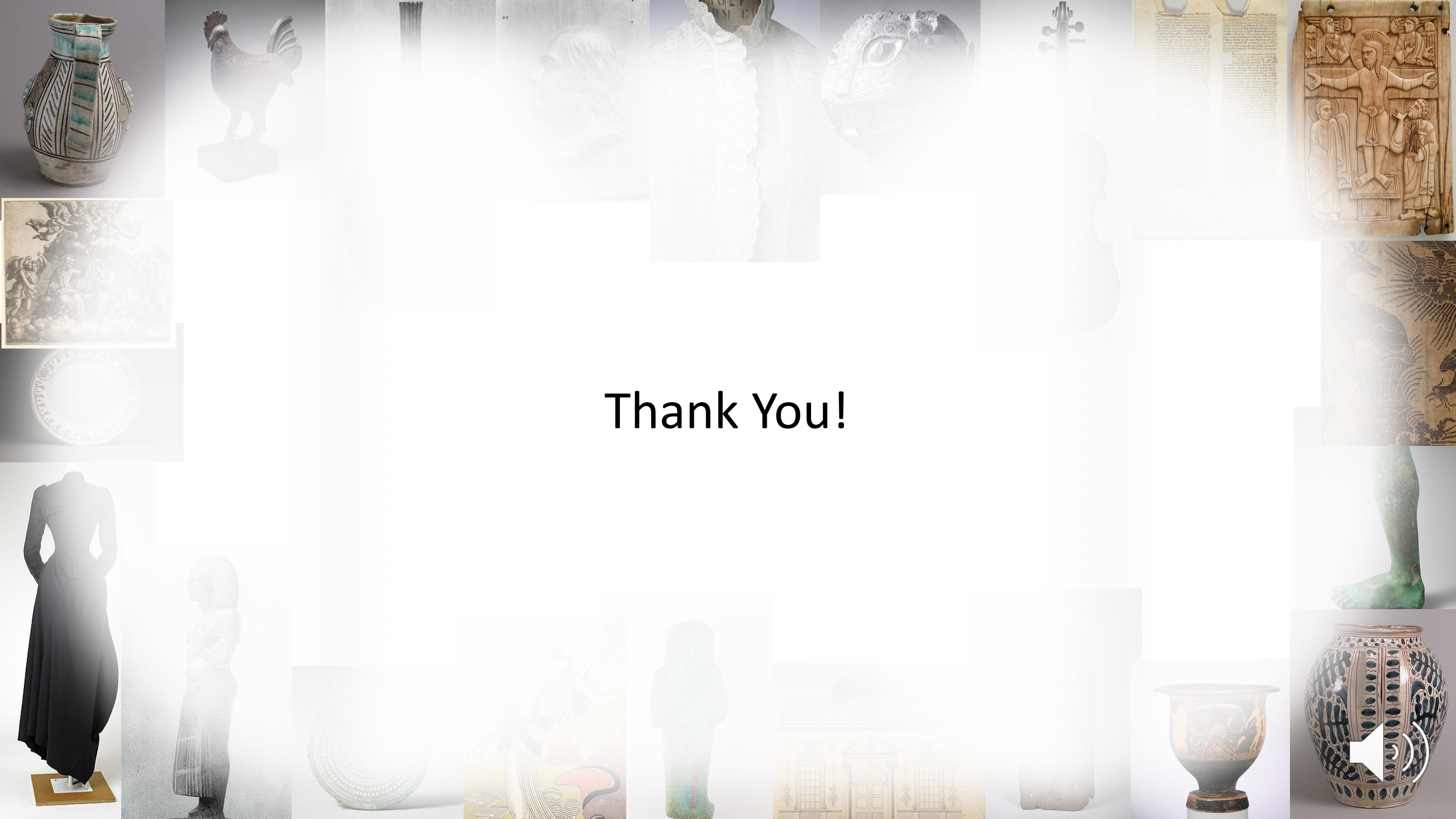
P

$P = NP$   
 $\approx$  NP-Complete

Complexity

$P \neq NP$

$P = NP$



Thank You!