Rank-based ordinal classification

Joan Serrat, Idoia Ruiz
Computer Vision Center, Universitat Autònoma de Barcelona

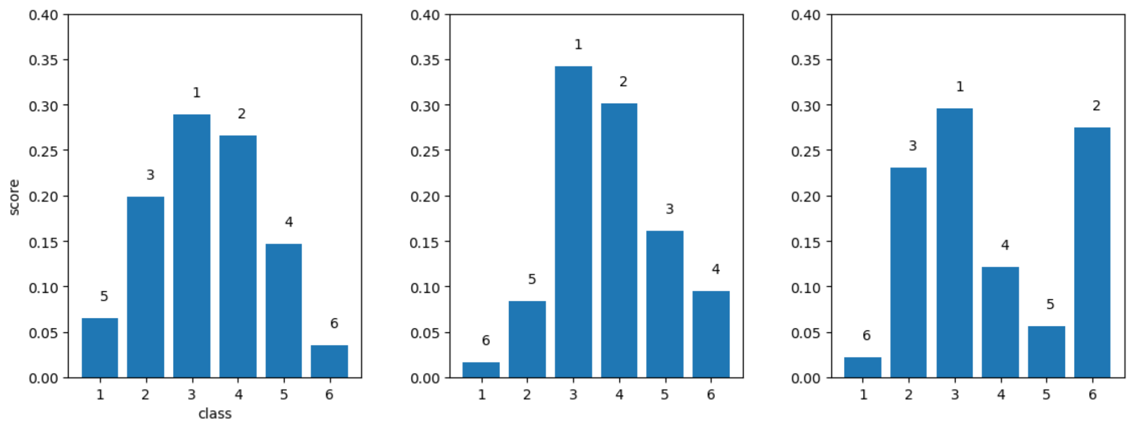
ORDINAL CLASSIFICATION

- Ordinal classification: Categories follow a relative order.
- Distance among categories can be unknown, *e.g.* What is the distance between a *professional* photo and a *flawed* one? Or between predicting a building destruction degree as *destroyed* or *severe damage*?

Idea

- Predict a ranking of all the ordinal classes, from most to least probable
- We propose a new ordinal classification loss that does not need to define a distance between classes. It compares predicted rankings to the groundtruth
- We enforce both the accuracy and consistency of prediction: the order of the classes must follow a unimodal distribution, which mode is the ground truth class.

Groundtruth class: 3



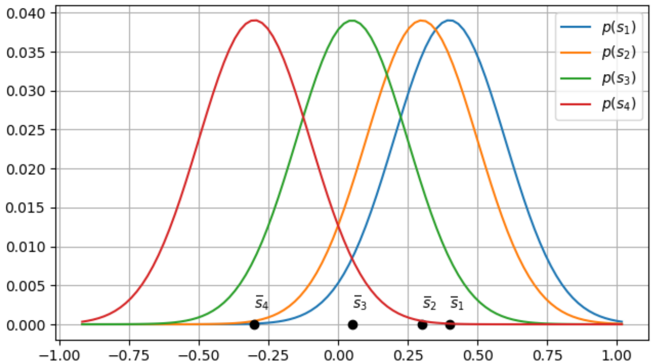
Predicted ranking	[3,4,2,5,1,6] consistent	[3,4,5,6,2,1] consistent	[3,2,3,5,1,6] not consistent

Rank-based ordinal classification

Joan Serrat, Idoia Ruiz
Computer Vision Center, Universitat Autònoma de Barcelona

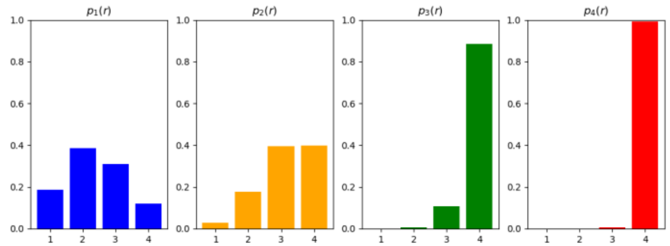
METHOD

From logits to rank probabilities



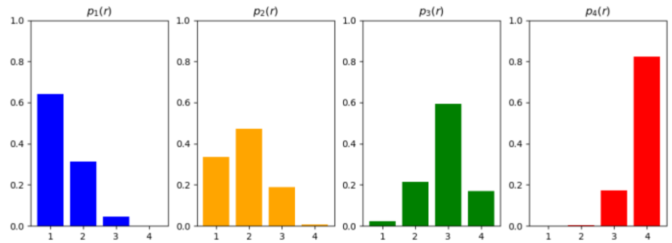
Let C = 4 be the number of classes and $\bar{s}_i,\ i \in [1,4]$ their scores, approximated as the mean of a Gaussian distribution with $\sigma = 0.2$

$$p(s_i) = \mathcal{N}(s_i | \bar{s}_i,\ \sigma^2)$$



$$P = [\, p_j(r)\,],\ j, r \in [1, C]$$

Probability for each scores to have each possible rank.



Same after 50 iterations of Sinkhorn transform.

- **One configuration loss**
  Valid ground truth rankings:
  $c_l = [l, l+1, l-1, l+2, l-2 \ldots]$
  where *l* is the true label

$$L_{\text{oc}}(P, l) = \sum_{r=1}^{C} \text{NLL}(\ [\, p_j(r)\,]_{j=1 \ldots C}, \text{OneHot}(c_l[r])\ )$$

- **All configurations loss**
  Valid ground truth rankings: The true label *l* is the mode and the sourronding classes have decreasing scores as in a Gaussian distribution centered at the true class.

$$L_{\text{ac}}(P, l) = \min_{v \in V(l)} L_{\text{oc}}(P, v)$$

- **Valid pairs loss**
  Similar to *All configurations* but computation is faster.

$$L_{\text{vp}}(P, l) = -\sum_{\substack{i<j\leq l, a>b \\ l \leq i < j, a<b}} \log p_i(a)\, p_j(b)$$

[1] M. Taylor, J. Guiver, S. Robertson, and T. Minka. Softrank: Optimising non-smooth rank metrics, in First ACM Int. Conf. on Web Search and Data Mining (WSDM), 2008.

Rank-based ordinal classification

Joan Serrat, Idoia Ruiz
Computer Vision Center, Universitat Autònoma de Barcelona

AESTHETICS RATING: SCHIFFANELLA'S IMAGEAESTHETICS

unacceptable 0.3% flawed 4.3% ordinary 72.4% professional 22.0% exceptional 1.0%

Category	EMD Acc.	EMD MAE	SORD Acc.	SORD MAE	Valid pairs Acc.	Valid pairs MAE
Nature	71.96*	0.342*	73.59†	0.271†	**74.77**	**0.261**
	72.06*	0.317*	71.04*	0.381*	**74.95**	**0.260**
Animals	66.98*	0.408*	**70.29**†	**0.308**†	69.32	0.318
	67.17*	0.405*	64.76*	0.555*	**70.07**	**0.310**
Urban	70.89*	0.342*	**73.25**†	**0.276**†	72.98	0.281
	70.64*	0.303*	67.75*	0.498*	**73.41**	**0.276**
People	67.97*	0.429*	70.59†	**0.309**†	**70.73**	**0.309**
	67.04*	0.421*	65.50*	0.571*	**70.79**	**0.307**

Top rows VGG16, bottom rows ResNet18. $\star$ as computed by our implementation. $\dagger$ as reported in papers.
CNN-POR not included because SORD is better in all categories.

INFORMATION RETRIEVAL: MSRA-MM

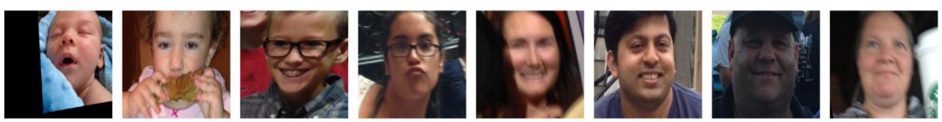
very relevant 35.5% relevant 42% irrelevant 22.5%

Samples for the query "Beach"

Query	CNN-POR Acc.	CNN-POR MAE	One config. Acc.	One config. MAE	All configs. Acc.	All configs. MAE	Valid pairs Acc.	Valid pairs MAE
Baby	50.00	0.636	51.26	0.590	**51.51**	0.592	51.35	**0.578**
Cat	52.89	0.598	54.07	0.534	**54.82**	0.536	54.09	**0.530**
Beach	51.11	0.596	**55.30**	0.496	54.85	0.503	55.27	**0.489**
Fish	66.33	0.355	67.48	0.337	66.63	0.337	**68.80**	**0.324**

LeNet and mean of 3 runs like in CNN-POR

AGE ESTIMATION: ADIENCE

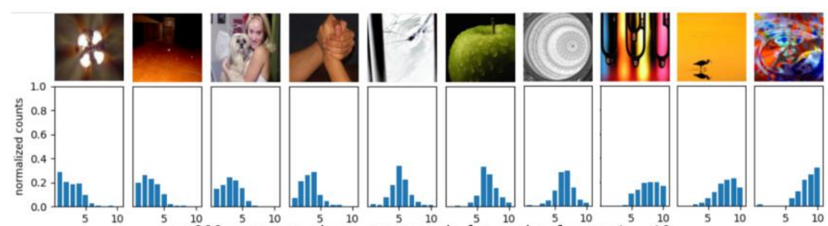


0-2 15% | 4-6 13% | 8-12 13% | 15-20 10% | 25-32 27% | 38-43 13% | 48-53 5% | 60+ 5%

	EMD	SORD	CNN-POR	One config	All configs.
Accuracy	**62.2** † —	59.6 $\pm$ 3.6 †	57.4 $\pm$ 5.8 †	55.3 $\pm$ 4.4	55.2 $\pm$ 3.7
	53.0 $\pm$ 5.3 *	48.8 $\pm$ 6.9 *	—	59.1 $\pm$ 5.2	59.0 $\pm$ 3.7
MAE	—	**0.49** $\pm$ 0.05 †	0.55 $\pm$ 0.08 †	0.57 $\pm$ 0.05	0.56 $\pm$ 0.05
	0.76 $\pm$ 0.09 *	1.31 $\pm$ 0.21 *	—	**0.49** $\pm$ 0.06	**0.49** $\pm$ 0.05

$\dagger$ as reported in these papers, single run of the experiment.
— not reported or implemented.
Odd rows VGG16, even rows ResNet18.

MOST VOTED CLASS PREDICTION: AVA



$\approx$ 200 votes per photo, one sample for mode of votes 1 . . . 10

New task analogous to mean score regression : predict the the most voted score.

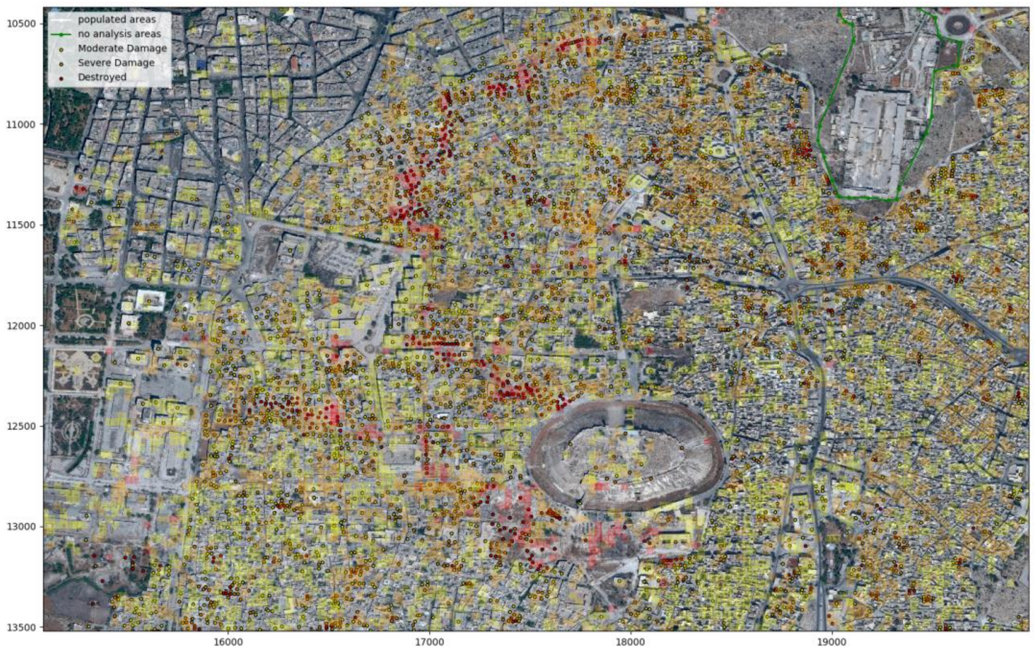
		SOTA	One config.
Mean score prediction	ρ_s	0.64	0.58
Most voted score prediction	Acc. %	—	63.15
	MAE	—	0.41

Rank-based ordinal classification

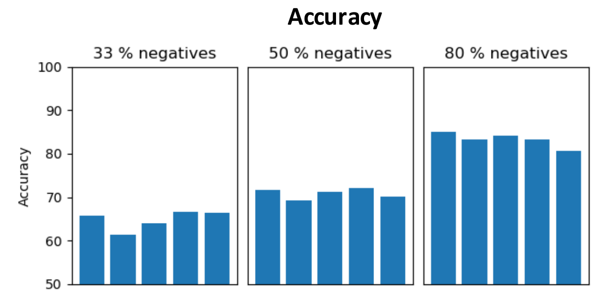
Joan Serrat, Idoia Ruiz
Computer Vision Center, Universitat Autònoma de Barcelona

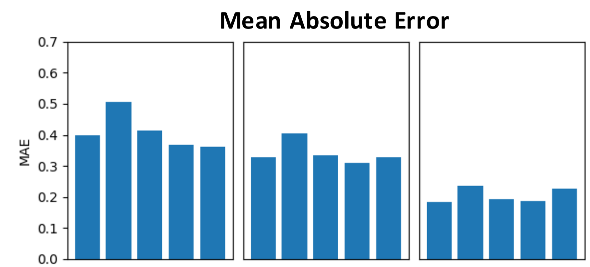
BUILDING DAMAGE ASSESSMENT

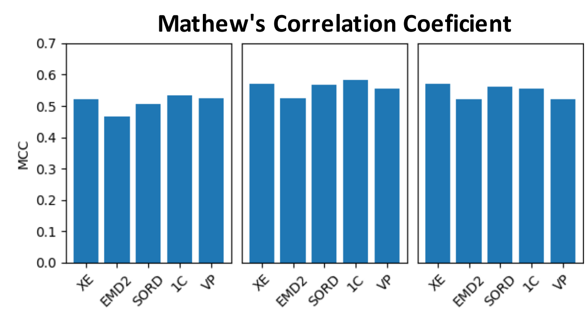
No visible damage | Moderate damage | Severe damage | Destroyed



Dots: Groundtruth. **Shaded squares:** Prediction (moderate damage, severe damage, destroyed)







Rank-based ordinal classification

Joan Serrat, Idoia Ruiz
Computer Vision Center, Universitat Autònoma de Barcelona

SUMMARY

- New method for ordinal classification that does not depend on the difference/distance between class labels.
- Three loss functions that compare groundtruth and predicted rankings, and enforce consistency in the prediction.
- We compare our method with SOTA on three different datasets, achieving similar or better results.
- We tackle a new task on image aesthetics assessment, namely, the prediction of the most voted class.
- We present results on a last application, building damage assessment from remote sensing images.