Deep Ordinal Regression with Label Diversity

Axel Berg[1,2], Magnus Oskarsson[2], Mark O'Connor[1]

1 arm

2 Lund University

Contributions

- ✓ We introduce the concept of label diversity
- ✓ We provide implementation examples
- ✓ We show state-of-the art results on several benchmarks

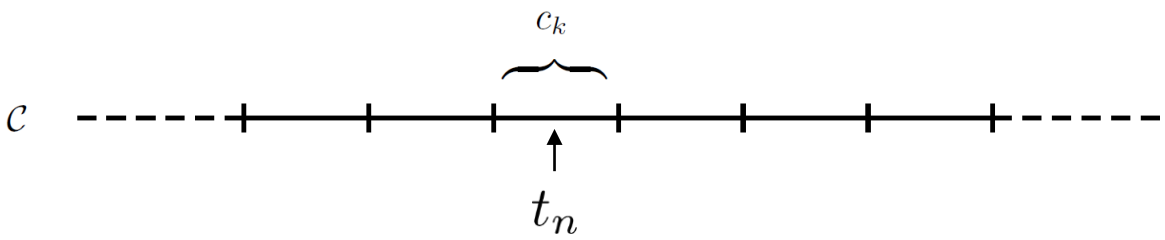
Regression via Classification (RvC)

- Inputs $x_n \in \mathbb{R}^{w \times h \times 3}$, targets $t_n \in \mathbb{R}$

- Direct regression $\min_{\theta} ||f_\theta(x_n) - t_n||^2$

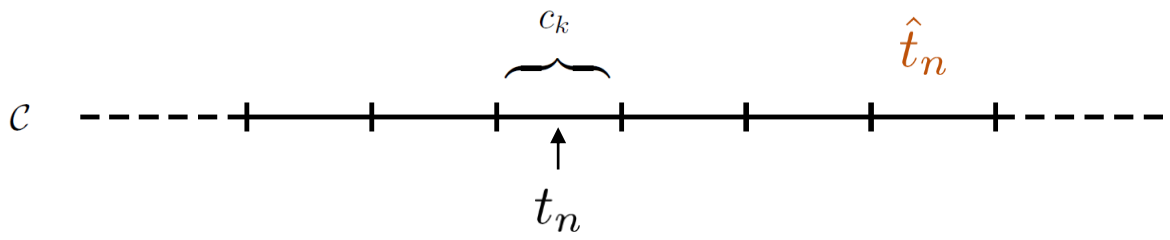
- RvC: define classes $\mathcal{C} = \{c_k\}_{k=1}^{K}$ and a distribution $q_n(c_k) = \begin{cases} 1, & t_n \in c_k, \\ 0, & \text{otherwise} \end{cases}$

Maximize cross entropy: $\max_{\theta} \sum_{k=1}^{K} q_n(c_k) p_\theta(c_k|x_n)$



Problems with RvC

- Baseline RvC does not exploit ordinal information
- How to choose the optimal binning of the target?



Proposed Solution:

✓ Introduce multiple sets of classes $\mathcal{D}^m = \{d_l^m\}_{l=1}^{L_m}$ for $m = 1, \dots, M$

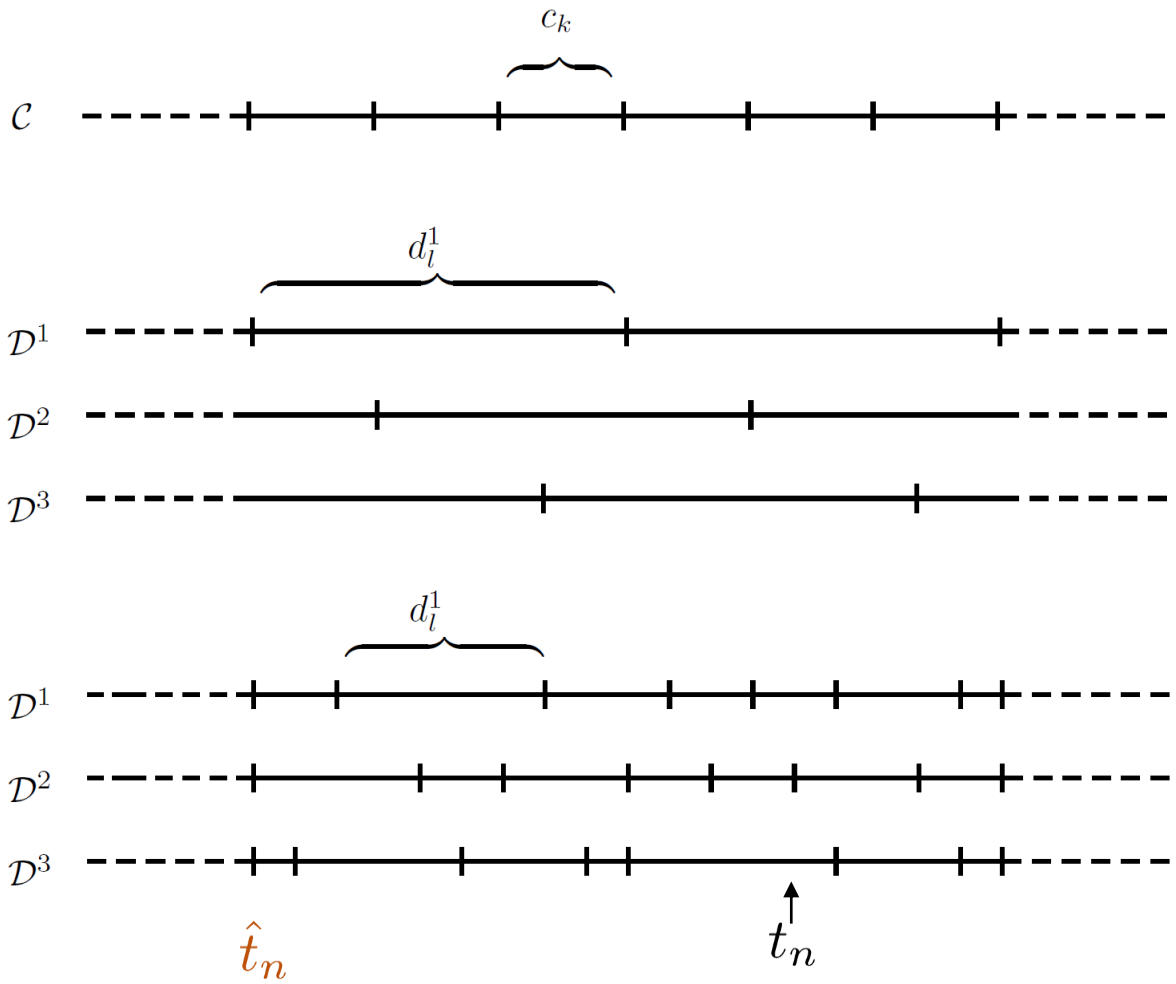
✓This induces diversity in the target representation

Label Diversity

Baseline RvC

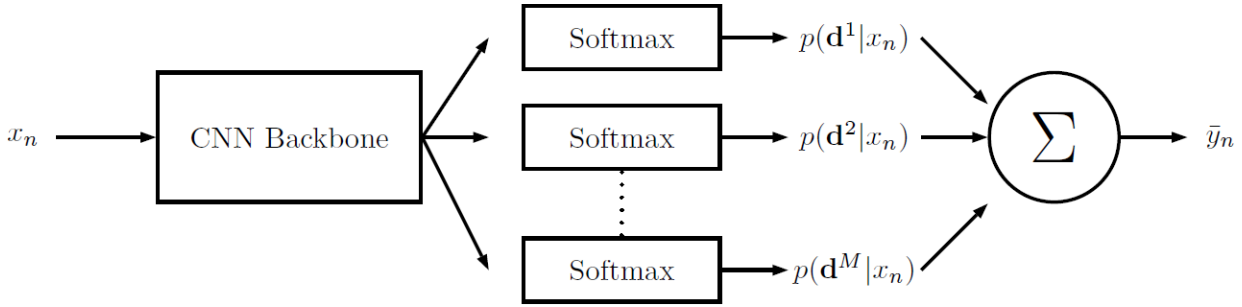
Equal Width

Randomized Bins



$$q_n(d_l^m) = \begin{cases} 1, & t_n \in d_l^m \\ 0, & \text{otherwise} \end{cases}$$

Implementation



$$\hat{y}_{m,n} = \sum_{l=1}^{L_m} w_l^m p(d_l^m|x_n)$$

Inference

$$\bar{y}_n = \frac{1}{M} \sum_{m=1}^{M} \hat{y}_{m,n}$$

$$(\bar{y}_n - t_n)^2 = \frac{1}{M} \sum_{m=1}^{M} (\hat{y}_{n,m} - t_n)^2 - \frac{1}{M} \sum_{m=1}^{M} (\hat{y}_{n,m} - \bar{y}_n)^2$$

Age Estimation on UTKFace

- Contains ~10 000 training images and ~5000 test images
- $t_n \in [21, 60]$ years

Age Estimation on UTKFace

- Pre-trained ResNet50 backbone
- Fine-tune with Adam optimizer for 30 epochs

Method	**SOTA**	**Direct (Ours)**	**RvC (Ours)**	**Equal Width (Ours)**	**Randomized Bins (Ours)**
Mean Average Error [Years]	4.65 ± 0.02	4.60 ± 0.02	4.71 ± 0.03	4.58 ± 0.03	**4.55 ± 0.04**

Age Estimation on UTKFace

- Pre-trained ResNet50 backbone
- Fine-tune with Adam optimizer for 30 epochs

Method	**SOTA**[1]	**Direct (Ours)**	**RvC (Ours)**	**Equal Width (Ours)**	**Randomized Bins (Ours)**
Mean Average Error [Years]	4.65 ± 0.02	4.60 ± 0.02	4.71 ± 0.03	4.58 ± 0.03	$\mathbf{4.55 \pm 0.04}$

1 F. K. Gustafsson, M. Danelljan, G. Bhat, and T. B. Schön, "Dctd: Deep conditional target densities for accurate regression," *arXiv preprint arXiv:1909.12297*, 2019.

Age Estimation on UTKFace

- Pre-trained ResNet50 backbone
- Fine-tune with Adam optimizer for 30 epochs

Method	**SOTA**	**Direct (Ours)**	**RvC (Ours)**	**Equal Width (Ours)**	**Randomized Bins (Ours)**
Mean Average Error [Years]	4.65 ± 0.02	4.60 ± 0.02	4.71 ± 0.03	4.58 ± 0.03	$\mathbf{4.55 \pm 0.04}$

$$C = \{21, \dots, 60\}$$

Age Estimation on UTKFace

- Pre-trained ResNet50 backbone
- Fine-tune with Adam optimizer for 30 epochs

Method	**SOTA**	**Direct (Ours)**	**RvC (Ours)**	**Equal Width (Ours)**	**Randomized Bins (Ours)**
Mean Average Error [Years]	4.65 ± 0.02	4.60 ± 0.02	4.71 ± 0.03	4.58 ± 0.03	**4.55 ± 0.04**
				$L = 8, M = 5$	

Age Estimation on UTKFace

- Pre-trained ResNet50 backbone
- Fine-tune with Adam optimizer for 30 epochs

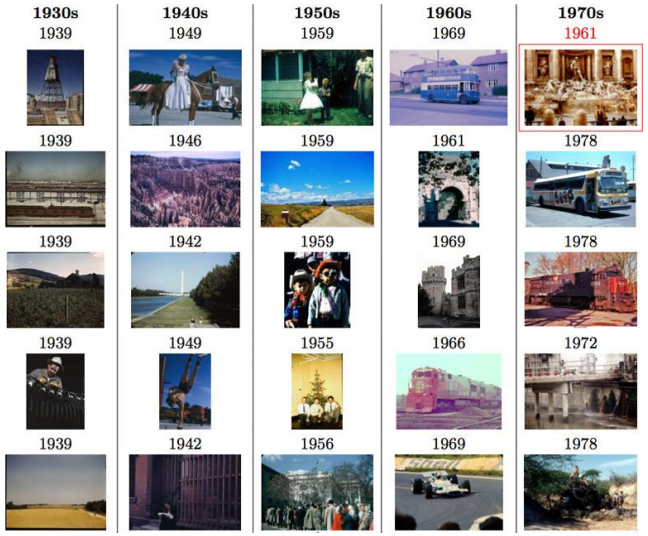
Method	SOTA	Direct (Ours)	RvC (Ours)	Equal Width (Ours)	Randomized Bins (Ours)
Mean Average Error [Years]	4.65 ± 0.02	4.60 ± 0.02	4.71 ± 0.03	4.58 ± 0.03	**4.55 ± 0.04**

$L = 10, M = 30$

More Results:

Historical Color Image Dating:

Method	CNNOr	PN	ELB	Direct (Ours)	RvC (Ours)	Label Diversity (Ours)
Acc (%)	41.56	56.67 ± 2.30	54.90 ± 2.40	46.9 ± 1.7	57.1 ± 1.9	**57.7 ± 2.0**
MAE	1.04	0.67 ± 0.05	**0.63 ± 0.04**	0.76 ± 0.01	0.72 ± 0.04	0.67 ± 0.03



Head Pose Estimation:

Method	Yang et al.	DCTD	Direct (Ours)	RvC (Ours)	Equal Width (Ours)	Randomized Bins (Ours)
Yaw	2.89	2.67 ± 0.08	2.64 ± 0.16	2.85 ± 0.12	2.62 ± 0.11	**2.52 ± 0.06**
Pitch	4.29	3.61 ± 0.12	**2.75 ± 0.05**	3.22 ± 0.09	3.12 ± 0.08	3.01 ± 0.10
Roll	3.60	2.75 ± 0.10	**2.24 ± 0.07**	2.48 ± 0.08	2.33 ± 0.09	2.34 ± 0.10
Average	3.60	3.01 ± 0.07	**2.54 ± 0.05**	2.85 ± 0.08	2.69 ± 0.04	2.63 ± 0.04

Conclusions

- Label diversity is a general framework for RvC problems
- Consistently improves on RvC baselines
- Future work: investigate potential theoretical guarantees

More details are provided in the paper

Code is available on Github

Thank you!