

# Multiset-Equivariant Set Prediction with Approximate Implicit Differentiation

Yan Zhang\*<sup>1</sup>, David W. Zhang\*<sup>2</sup>, Simon Lacoste-Julien<sup>1,3,4</sup>, Gertjan J. Burghouts<sup>5</sup>, Cees G. M. Snoek<sup>2</sup>

<sup>1</sup>Samsung – SAIT AI Lab, Montreal <sup>2</sup>University of Amsterdam <sup>3</sup>Mila, Université de Montreal <sup>4</sup>Canada CIFAR AI Chair <sup>5</sup>TNO \*equal contribution



UNIVERSITY OF AMSTERDAM

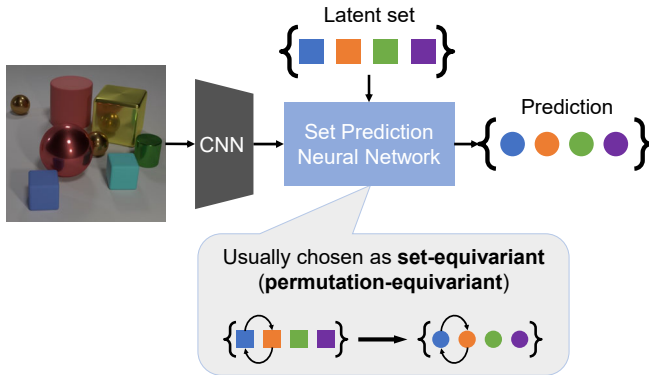


Mila

CIFAR

TNO

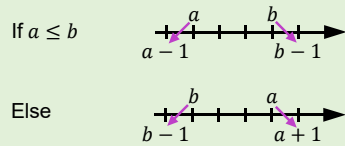
## Equivariance: Role in Set Prediction



## Limitation of Set-Equivariance

Q: Can set-equivariant neural networks represent **push\_apart**?

**Definition of push\_apart**  $([a, b])$ :  
Push smaller to the left and larger to the right.



Examples:

$$\text{push\_apart}([1, 2]) = [0, 3]$$

$$\text{push\_apart}([2, 1]) = [3, 0]$$

Looks like it's set-equivariant, *but what if*  $a = b$ ?

$$\text{push\_apart}([1, 1]) = [0, 2]$$

$$\text{push\_apart}([1, 1]) = [0, 2]$$

A: Set-equivariant functions **cannot represent** **push\_apart**!

**Multiset-equivariance** without set-equivariance is the proper permutation symmetry for set prediction and enables **separating equal elements**.

Video, code, pre-trained models and run statistics available at <https://github.com/davzha/multiset-equivariance>

## Multiset-Equivariance

Previous notion of permutation equivariance:  
A function  $f: \mathbb{R}^{n \times d_1} \rightarrow \mathbb{R}^{n \times d_2}$  is **set-equivariant** iff for any permutation matrix  $P$ :

$$f(PX) = Pf(X)$$

**Better:**  
A function  $f: \mathbb{R}^{n \times d_1} \rightarrow \mathbb{R}^{n \times d_2}$  is **multiset-equivariant** iff for any permutation matrix  $P_1$ , there exists  $P_2$  such that  $P_1X = P_2X$ :

$$f(P_1X) = P_2f(X)$$

## Implicit Deep Set Prediction Networks

$$\text{iDSPN}(z) = \arg \min_Y L(Y, z, \theta)$$

**Forward-pass:**

Find the best output set  $Y$ , by minimizing  $L(Y, z, \theta) = \|g(Y, \theta) - z\|$  via gradient descent.

**Backward-pass:**

Calculate gradients for  $\theta = (z, \theta)$  using (approximate) implicit differentiation. In detail, the optimality condition is:

$$\nabla_Y L(Y^*, \theta) = 0$$

We differentiate both sides to get a linear system  $HX = B$ :

$$\frac{\partial \nabla_Y L(Y^*, \theta)}{\partial Y^*} \frac{\partial Y^*}{\partial \theta} = - \frac{\partial \nabla_Y L(Y^*, \theta)}{\partial \theta}$$

Solving the linear system for  $X$  gives us the Jacobian for backprop.

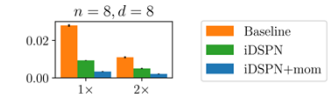
## Multiset-equivariance more powerful than set-equivariance

**Example:** Given  $[a, a, b, b, b]$ , number each letter separately:  
 $[(a, 0), (a, 1), (b, 0), (b, 1), (b, 2)]$ .

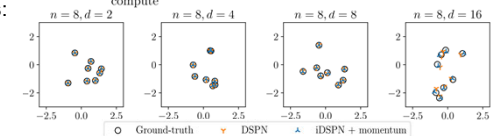
Model	Set-equivariant	Multiset-equivariant	Training samples		
			1x	10x	100x
DeepSets	✓	✓	0.0±0.0	0.0±0.0	0.0±0.0
Transformer without PE	✓	✓	0.0±0.0	0.0±0.0	0.0±0.0
Transformer with random PE	✗	✗	28.5±8.1	22.6±5.2	22.7±7.7
Transformer with PE	✗	✗	0.0±0.0	46.4±26.2	94.4±1.8
BiLSTM	✗	✗	0.0±0.0	25.1±20.2	97.9±2.1
DSPN	✗	✓	89.4±2.4	88.9±4.9	90.8±4.2
iDSPN (ours)	✗	✓	96.8±0.6	98.1±0.3	97.9±0.6

## Implicit DSPN greatly improves over baseline DSPN

**Tasks:** Auto-encode random  $d$ -dimensional point sets of size  $n$ .



Examples:



## Much more precise than SotA on CLEVR Object Prediction

Model	Image size	AP <sub>∞</sub>	AP <sub>1</sub>	AP <sub>0.5</sub>	AP <sub>0.25</sub>	AP <sub>0.125</sub>	AP <sub>0.0625</sub>
Slot MLP	128x128	19.8±1.6	1.4±0.3	0.3±0.2	0.0±0.0	0.0±0.0	—
DSPN	128x128	85.2±4.8	81.1±5.2	47.4±17.6	10.8±9.0	0.6±0.7	—
Slot Attention	128x128	94.3±1.1	86.7±1.4	56.0±3.6	10.8±1.7	0.9±0.2	—
Slot Attention*	128x128	96.4±0.5	93.6±0.8	80.4±2.1	26.5±2.7	2.6±0.3	—
Slot Attention†	128x128	90.6±1.8	89.1±2.1	84.4±2.5	50.3±3.8	7.9±1.0	—
iDSPN	128x128	98.8±0.5	98.5±0.6	98.2±0.6	95.8±0.7	76.9±2.5	32.3±3.9
iDSPN	256x256	99.4±0.3	99.2±0.4	99.0±0.5	97.8±0.7	86.9±1.0	47.2±3.8

Image	Ground-truth	iDSPN 10 iterations	iDSPN 20 iterations	iDSPN 40 iterations
	(-2.95 -2.56 0.70) large purple metal cylinder (-2.26 0.66 0.35) small cyan metal sphere (-2.13 1.62 0.35) small brown metal cylinder (-0.83 -1.80 0.35) small red rubber sphere (0.34 -2.86 0.35) small cyan rubber cylinder (0.37 2.25 0.70) large cyan rubber cube (1.25 -2.94 0.70) large yellow rubber cylinder (1.41 0.07 0.35) small purple metal cylinder (0.02 2.29 0.35) small brown metal cylinder (2.93 -2.16 0.70) large red metal cube	(-2.79 -2.19 0.64) d=0.41 large purple metal cylinder (-2.48 0.60 0.26) d=0.24 small cyan metal sphere (-1.46 2.06 -0.06) d=0.90 small brown metal cylinder (-0.69 -2.33 0.35) d=0.55 small yellow rubber sphere (0.73 0.96 0.44) d=1.36 small red rubber cube (0.55 -2.16 -0.49) d=1.59 large brown metal cylinder (0.59 0.13 -0.12) d=0.95 small purple metal cylinder (2.04 2.23 -0.04) d=0.40 small brown metal cylinder (3.15 -2.34 0.68) d=0.29 large red metal cube	(-2.85 -2.58 0.79) d=0.13 large purple metal cylinder (-2.32 0.41 0.41) d=0.26 small cyan metal sphere (-2.04 1.70 0.41) d=0.14 small brown metal cylinder (-0.81 -1.84 0.32) d=0.05 small red rubber sphere (-0.08 -2.73 0.43) d=0.30 small yellow rubber cylinder (1.04 -2.36 0.84) d=0.63 large yellow rubber cylinder (0.44 2.24 0.70) d=0.06 large cyan rubber cube (1.04 -2.06 0.84) d=0.14 small purple metal cylinder (1.09 0.05 0.46) d=0.39 small purple metal cylinder (1.97 2.21 0.46) d=0.11 small brown metal cylinder (2.96 -2.34 0.83) d=0.23 large red metal cube	(-2.89 -2.53 0.67) d=0.07 large purple metal cylinder (-2.31 0.45 0.33) d=0.21 small cyan metal sphere (-2.03 1.74 0.37) d=0.16 small brown metal cylinder (-0.81 -1.83 0.33) d=0.04 small red rubber sphere (-0.24 -2.86 0.33) d=0.11 small cyan rubber cylinder (1.30 -2.90 0.72) d=0.07 large yellow rubber cylinder (0.42 2.28 0.68) d=0.06 large cyan rubber cube (1.09 0.10 0.31) d=0.33 small purple metal cylinder (1.96 2.26 0.38) d=0.07 small brown metal cylinder (2.90 -2.20 0.70) d=0.05 large red metal cube