NNDL@UNIPD

ENSEMBLING CNN MODELS FOR 3D OBJECT CLASSIFICATION BASED ON VOXEL GRID REPRESENTATION

Authors

D'Antimo Simone[1] | simone.dantimo@studenti.unipd.it
Singh Harjot[1] | harjot.singh@studenti.unipd.it

[1]The University of Padua

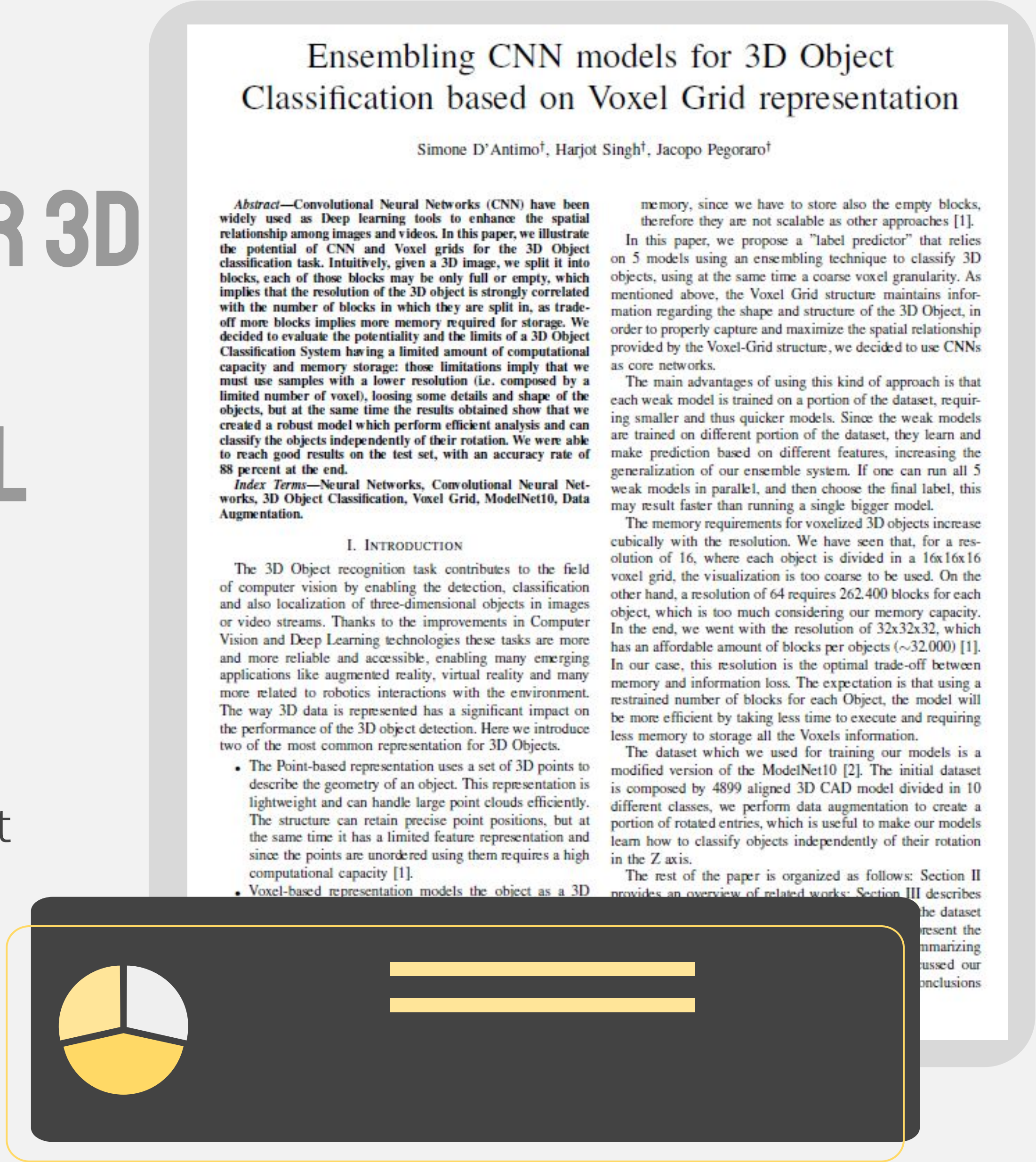


TABLE OF CONTENTS

THE CHALLENGE

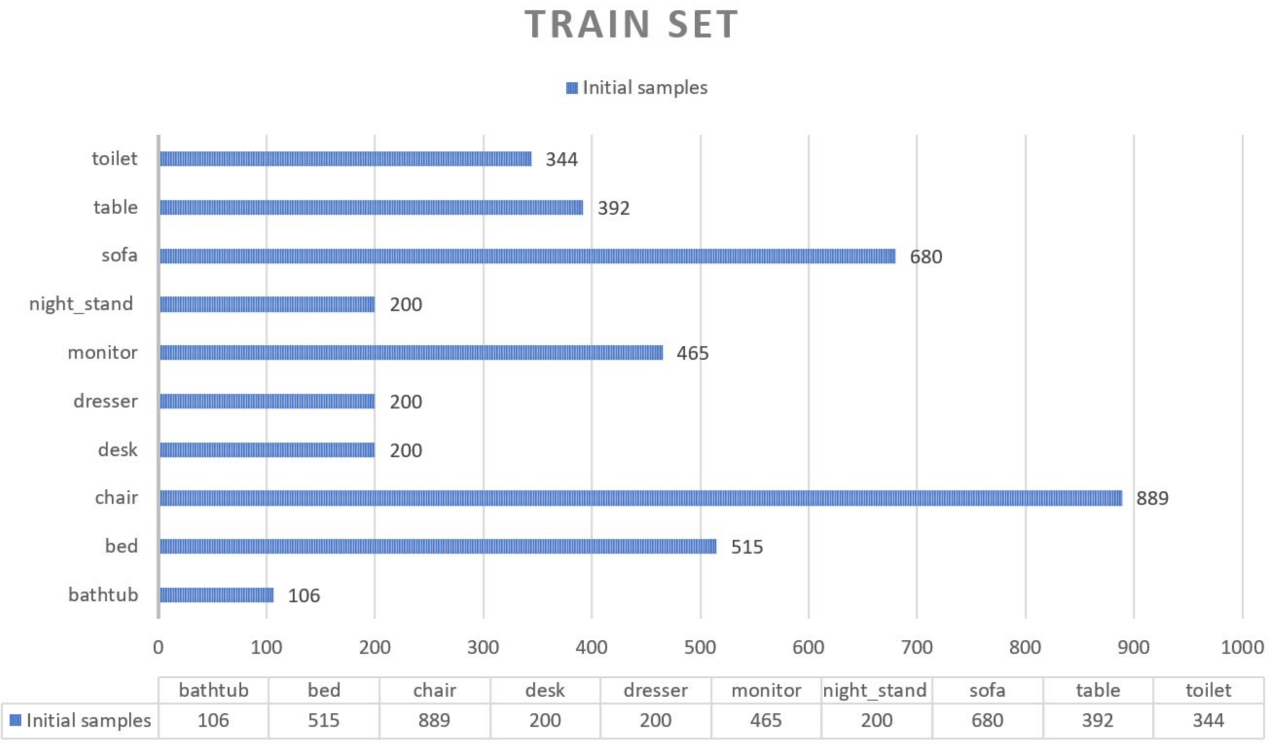
THE CHALLENGE

Our goal: Classifying 3D Object independently their rotation using CNN and Voxel Grid representation

Our data: ModelNet10,

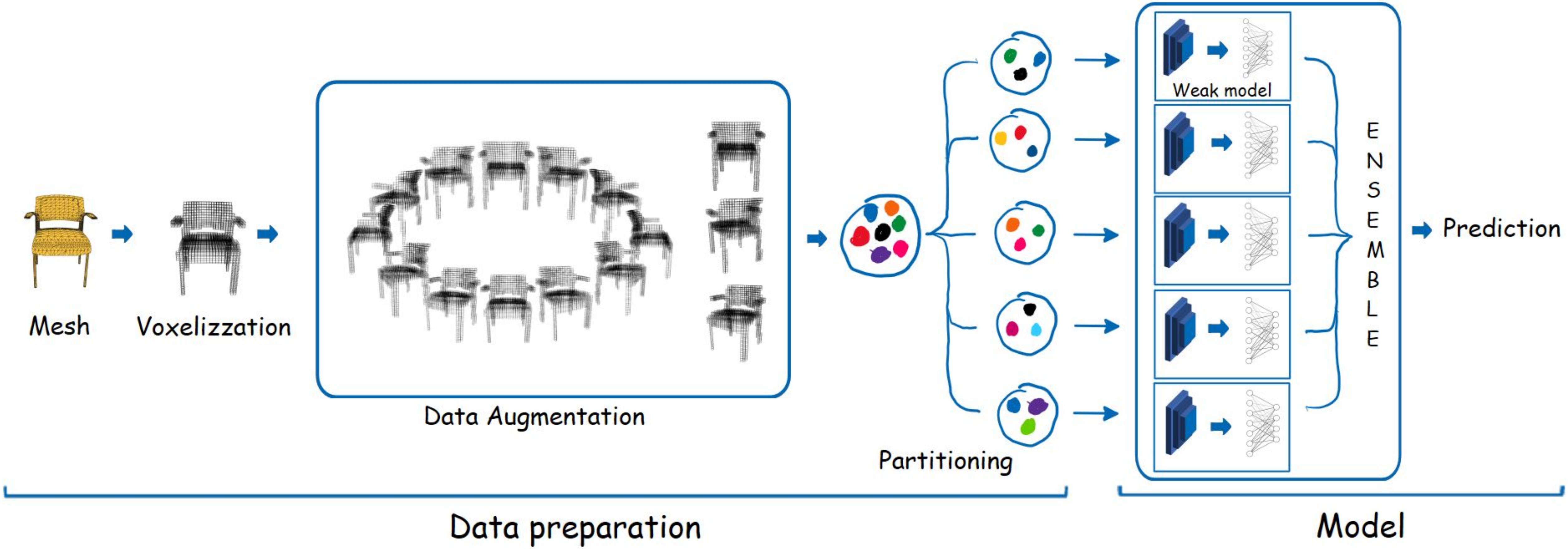
- 4899 3D CAD divided into 10 classes
- 80/20* = train/test* splitting
- orientation aligned

- CAD models
- Non i.i.d.
- Desired partition 80-10-10



OUR APPROACH

PROCESSING PIPELINE



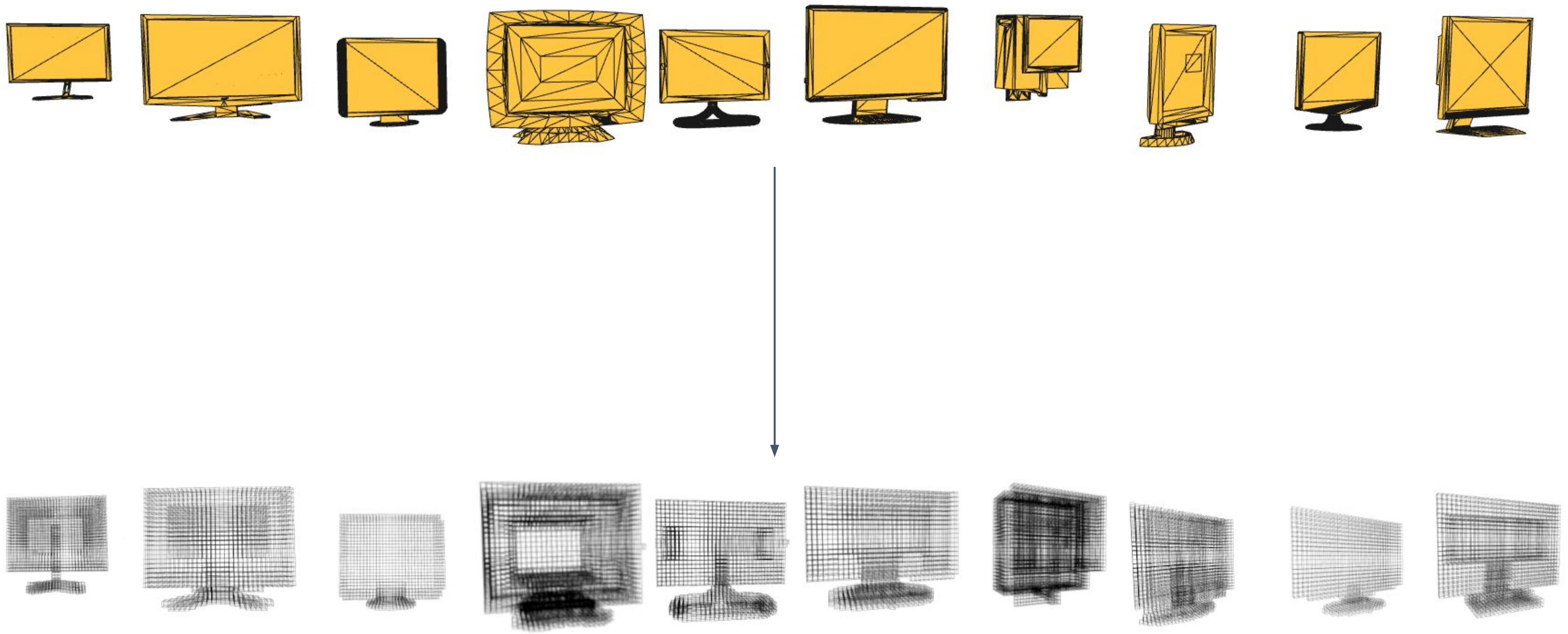
DATA PREPARATION

DATA PREPARATION

VOXELIZATION

1

ModelNet10
3D Cad Models



Voxel Grids

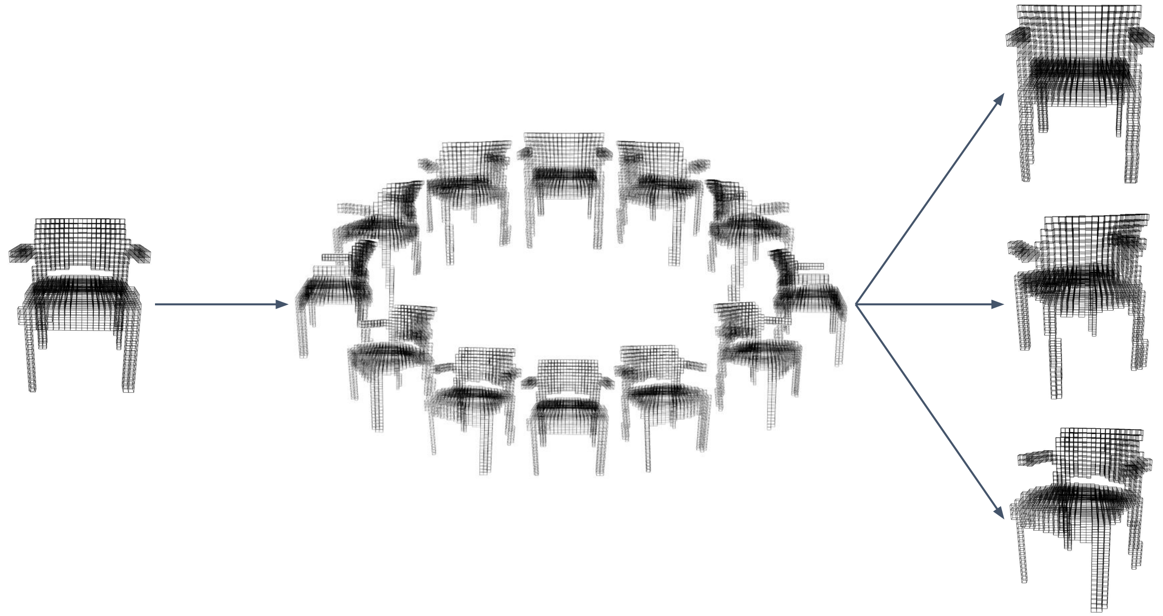
DATA PREPARATION

VOXELIZATION > **AUGMENTATION** >

1. **Focus** on making the **dataset IID**

2. **N Rotations,** depending on their initial distribution

Rotations	Classes
2	bed, chair, monitor, sofa
3	table, toilet
5	desk, dresser, night_stand
8	bathtub



DATA PREPARATION

VOXELIZATION > **AUGMENTATION** >

1. **Focus** on making the **dataset IID**
2. **N Rotations,** depending on their initial distribution

Rotations	Classes
2	bed, chair, monitor, sofa
3	table, toilet
5	desk, dresser, night_stand
8	bathtub

	bathtub	bed	chair	desk	dresser	monitor	night_stand	sofa	table	toilet
AFTER AUGMENTING	848	1000	1000	1000	1000	930	1000	1000	1176	1032
Initial samples	106	515	889	200	200	465	200	680	392	344

DATA PREPARATION

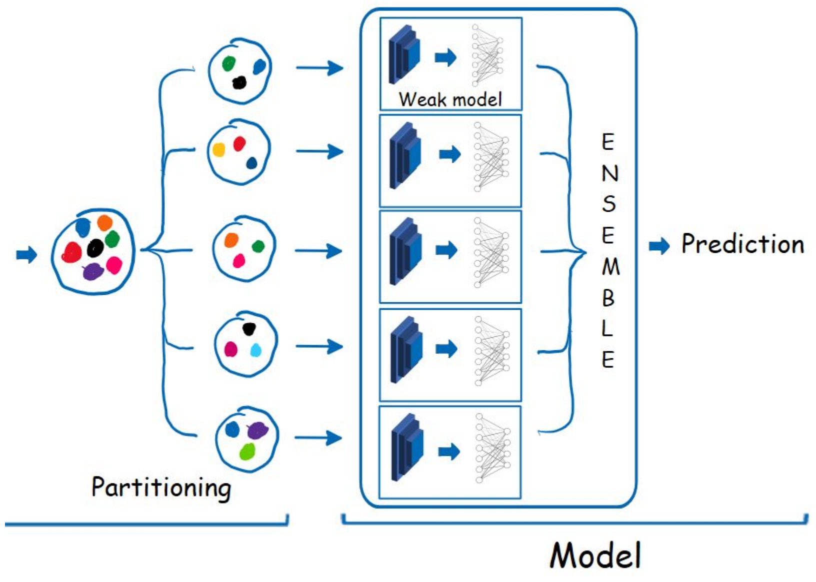
VOXELIZATION > AUGMENTATION > **PARTITIONING**

1. **Focus** on making the **dataset ready for the ENSEMBLE models**

2. Each weak model should learn
   - well on their part of the dataset
   - some information on data of other models.

TABLE 2: Our Bagging datasets compositions

Dataset	Chunk's subsets	Copied Subset
1	01, 02, 03, 04, 05	10, 15, 20, 25 — 24
2	06, 07, 08, 09, 10	01, 11, 16, 21 — 05
3	11, 12, 13, 14, 15	02, 07, 17, 22 — 06
4	16, 17, 18, 19, 20	03, 08, 13, 23 — 12
5	21, 22, 23, 24, 25	04, 09, 14, 19 — 18

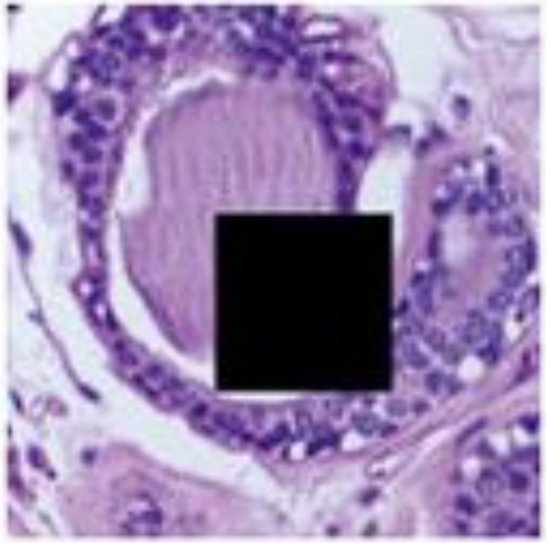


DATA PREPARATION

VOXELIZATION > AUGMENTATION > PARTITIONING > **CUTOUT**

1 **Masking Out** a random portion of the input data

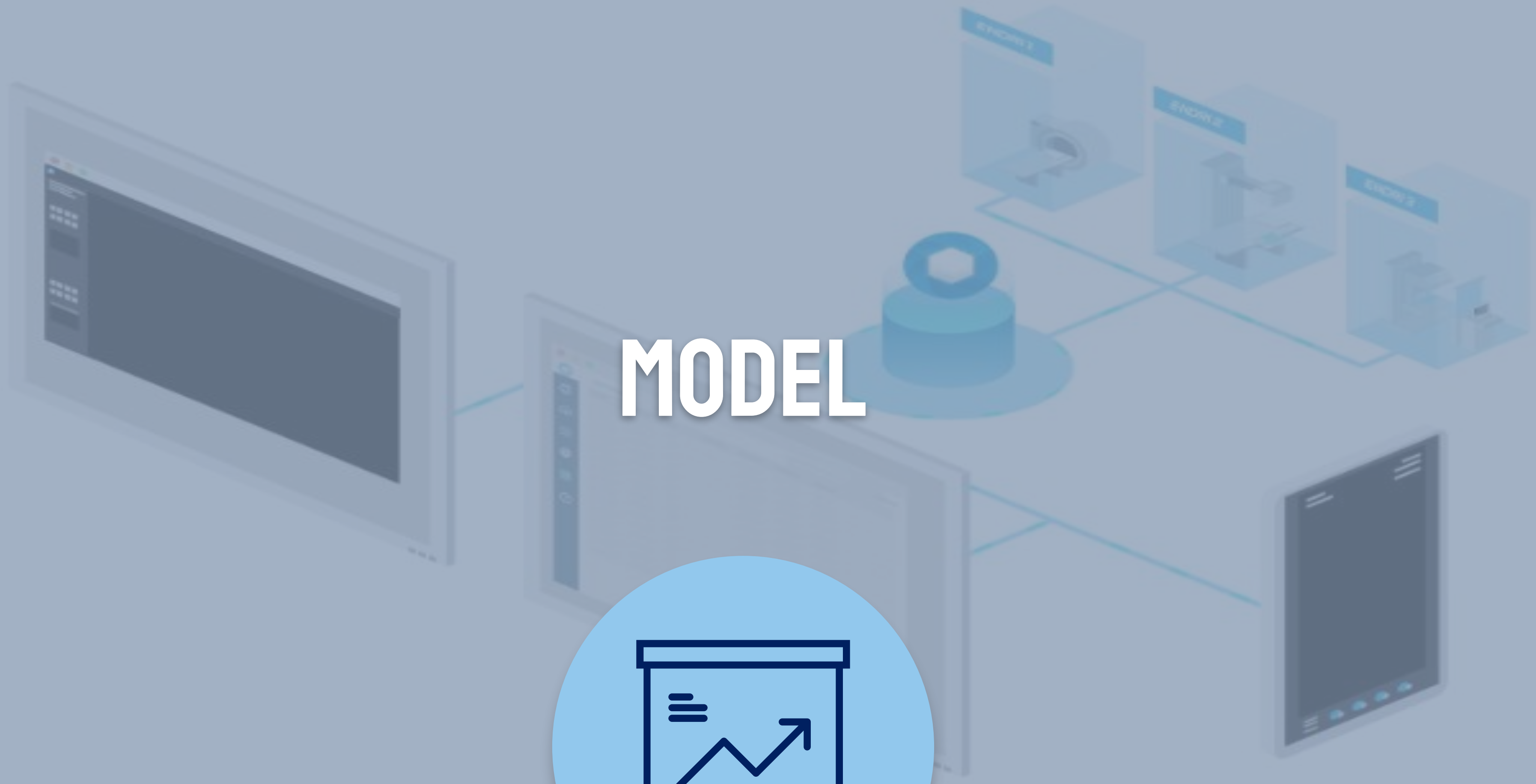
- Improve robustness
- Improve generalization



Cutout

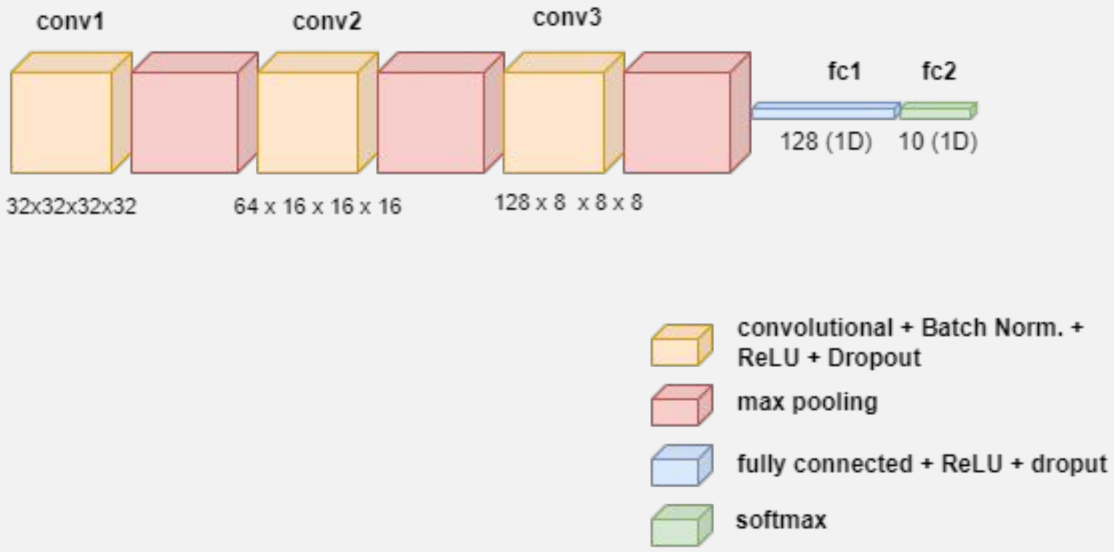
Reference: [Research Gate]

MODEL



THE WEAK MODELS STRUCTURE

Layer (type)	Output Shape	Param #
Conv3d-1	[-1, 32, 32, 32, 32]	896
BatchNorm3d-2	[-1, 32, 32, 32, 32]	64
ReLU-3	[-1, 32, 32, 32, 32]	0
Dropout-4	[-1, 32, 32, 32, 32]	0
MaxPool3d-5	[-1, 32, 16, 16, 16]	0
Conv3d-6	[-1, 64, 16, 16, 16]	55,360
BatchNorm3d-7	[-1, 64, 16, 16, 16]	128
ReLU-8	[-1, 64, 16, 16, 16]	0
Dropout-9	[-1, 64, 16, 16, 16]	0
MaxPool3d-10	[-1, 64, 8, 8, 8]	0
Conv3d-11	[-1, 128, 8, 8, 8]	221,312
BatchNorm3d-12	[-1, 128, 8, 8, 8]	256
ReLU-13	[-1, 128, 8, 8, 8]	0
Dropout-14	[-1, 128, 8, 8, 8]	0
MaxPool3d-15	[-1, 128, 4, 4, 4]	0
Linear-16	[-1, 128]	1,048,704
ReLU-17	[-1, 128]	0
Dropout-18	[-1, 128]	0
Linear-19	[-1, 10]	1,290



!! Represent in 3D for model flow, but 3DConv layer has 4 dimension

TRAINING & EXPERIMENTAL RESULTS

TRAINING WEAK MODELS

Epochs per model:100

- Converge quickly <50 Epochs

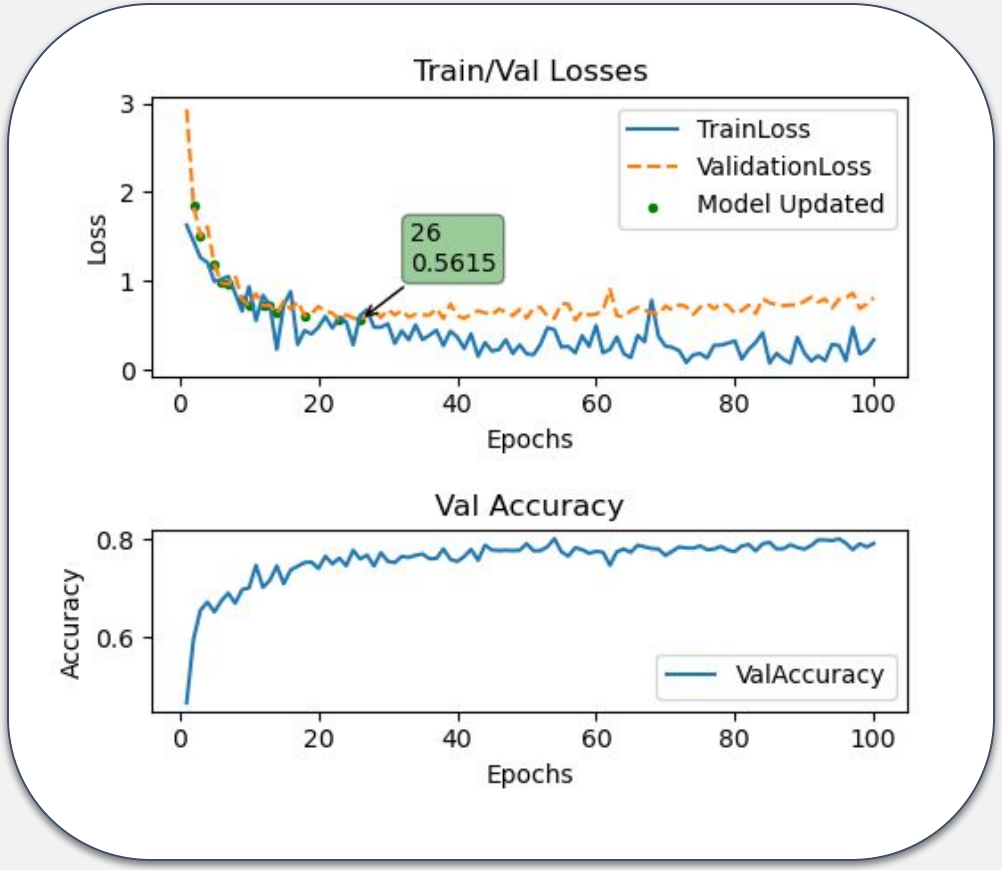
Learning rate: 1e-3

Optimizer: ADAM

Loss: Cross Entropy

Max Pooling Kernel: 2x2x2

Max Pooling Stride: 2x2x2



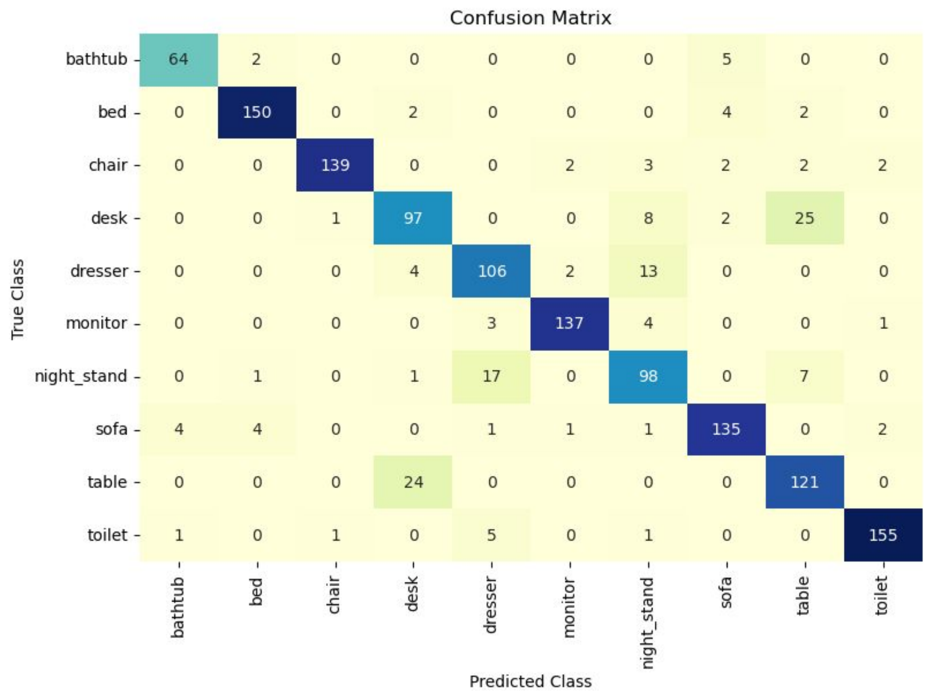
RESULTS ON VALIDATION SET

	CLASS	M0	M1	M2	M3	M4	EN
0	bathtub	0.98	0.98	0.94	0.97	0.93	0.98
1	bed	0.92	0.90	0.92	0.96	0.94	0.94
2	chair	0.96	0.91	0.98	0.94	0.96	0.97
3	desk	0.62	0.67	**0.80**	0.75	0.65	0.78
4	dresser	0.74	0.73	0.71	0.83	0.74	0.77
5	monitor	0.94	0.97	0.97	0.96	0.97	0.96
6	night_stand	0.71	0.75	0.75	0.81	0.67	0.75
7	sofa	0.88	0.76	0.81	0.87	0.90	0.88
8	table	0.67	0.77	0.81	0.87	0.90	0.88
9	toilet	**0.99**	0.90	0.95	0.91	0.98	0.97

RESULTS ON TEST SET

Ensembled model results:

- Good diagonal Confusion matrix
- Errors are "Symmetric"



Confusion matrix of the ensemble model

RESULTS ON TEST SET

		ACCURACY
1	M0	0.83
2	M1	0.83
3	M2	0.85
4	M3	0.87
5	M4	0.86
6	EN	**0.88**

BASELINE ACCURACY: 0.90

ShapeNet: 0.83
DeepPano 0.85
VoxNet 0.92
Orion 0.938

CONCLUSIONS AND FUTURE WORK

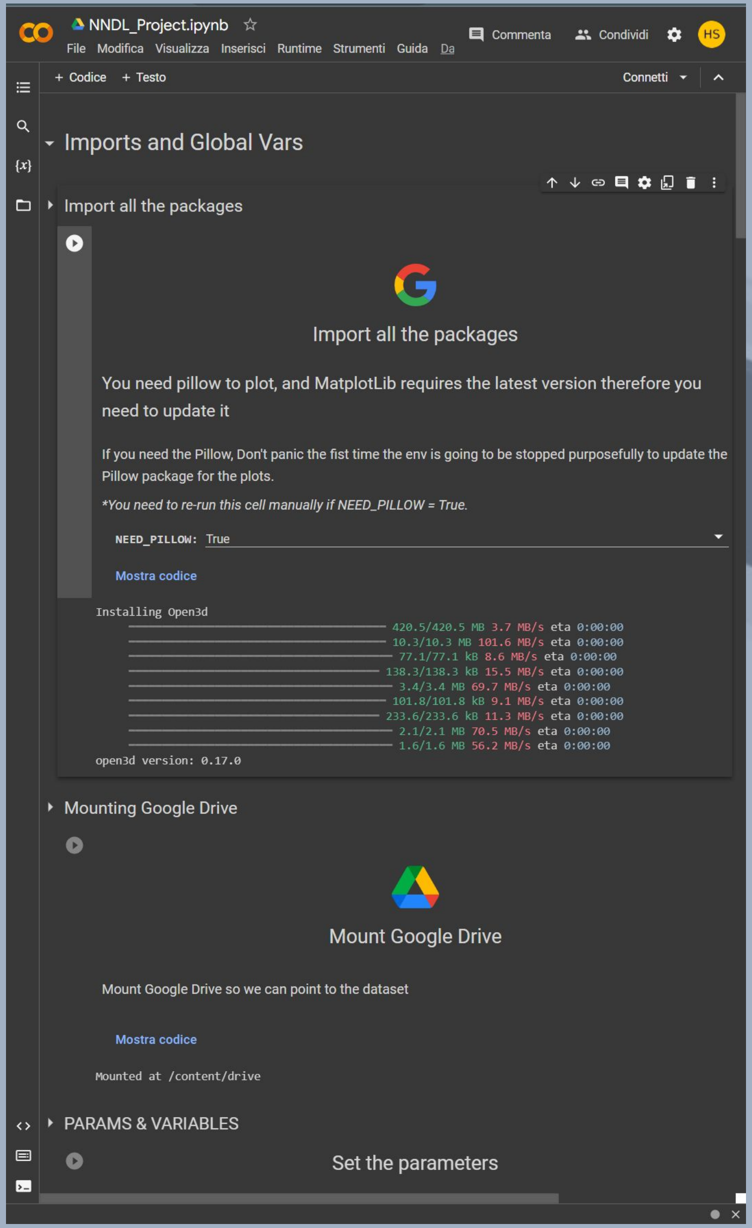
CONCLUSIONS

- Model has good accuracy
- Dynamic, can be adapted on the computer in which is run

FUTURE WORK

- Adding more data and Models
- Implements a way to capture long-range context information

END THANK YOU



NOTEBOOK DEMO