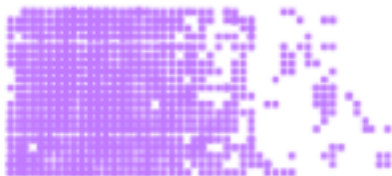


Sparse Submanifold Neural Network



















~

2

0

50x50

NO



W

1

M

2

MS

3

11

4

15

5

16

6













































Encoder











此
後
不
復
有
此
事
矣

此
後
不
復
有
此
事
矣



































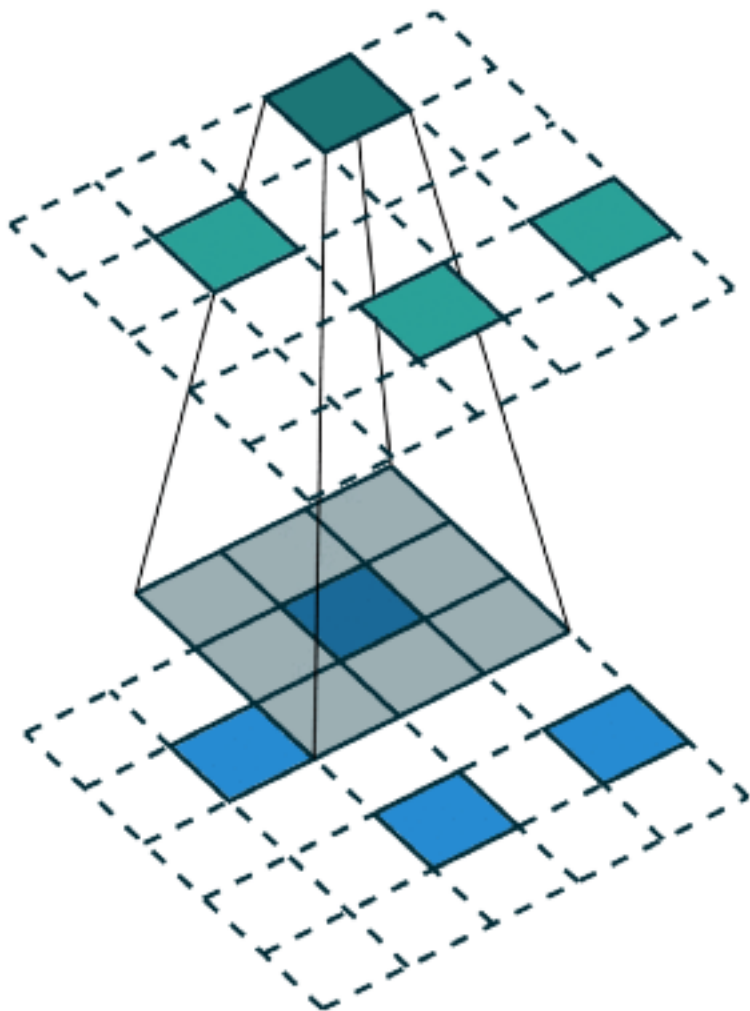






Pre-Training









1. *first*

768 is at its



16 or 16

1 vector embedding































SCNN reduces the spatial coordinates while increasing the feature dimensions.

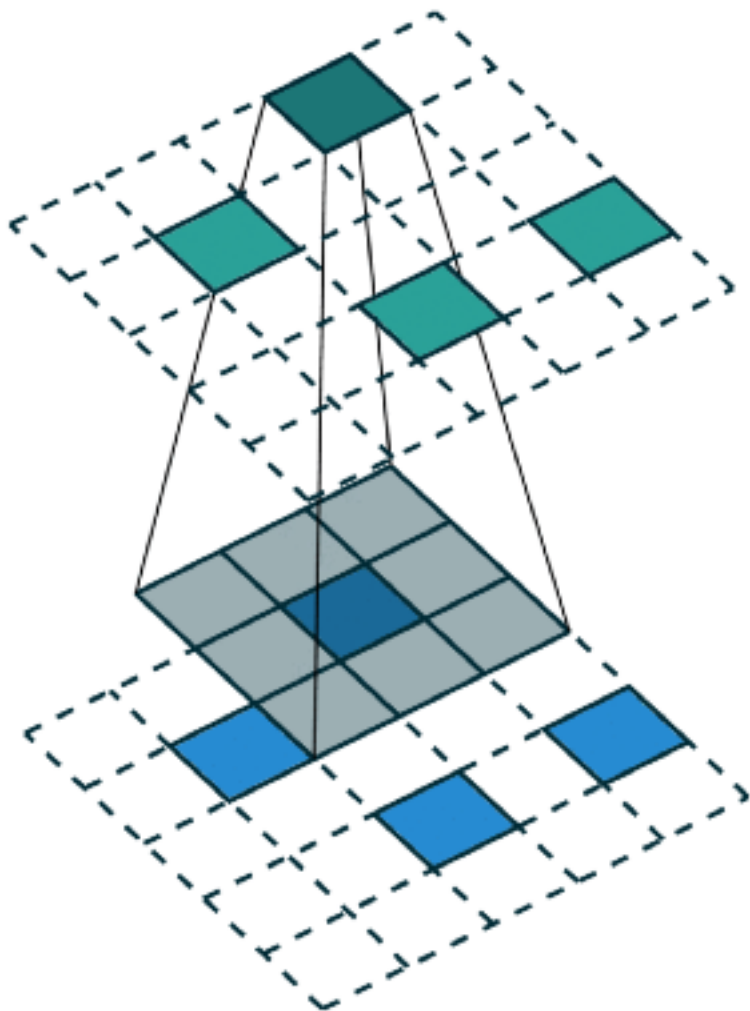
It performs convolution only on active voxels, effectively ignoring the vast empty regions typical in detector data.

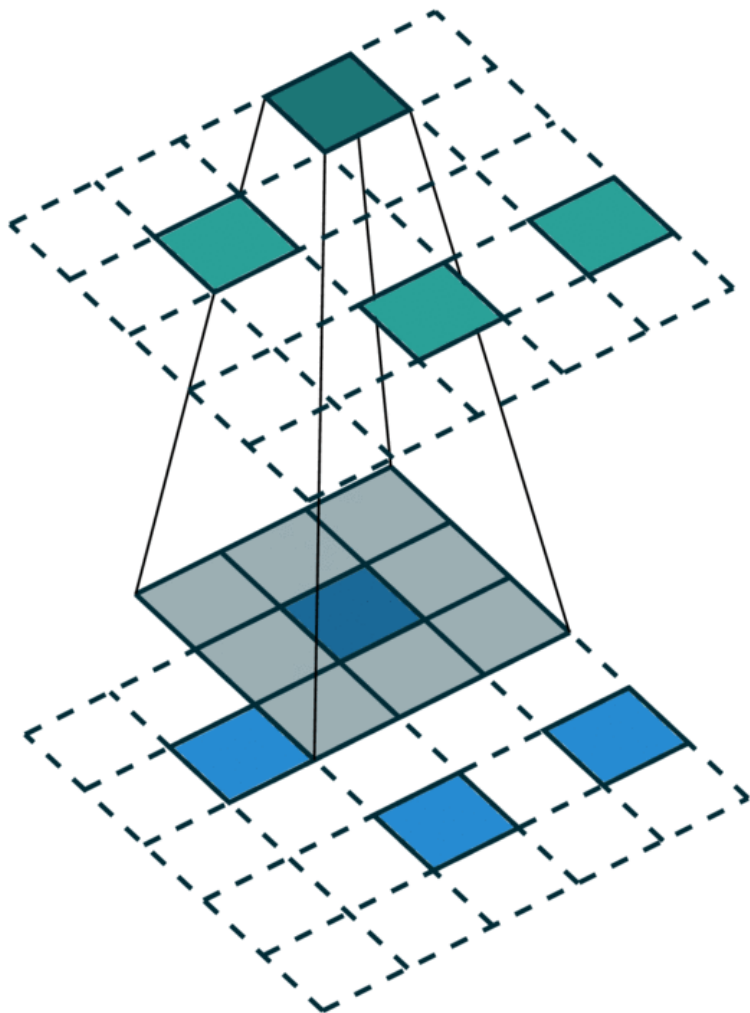
SCNN

SCNN: Creates a vector embedding for every patch in the input

3

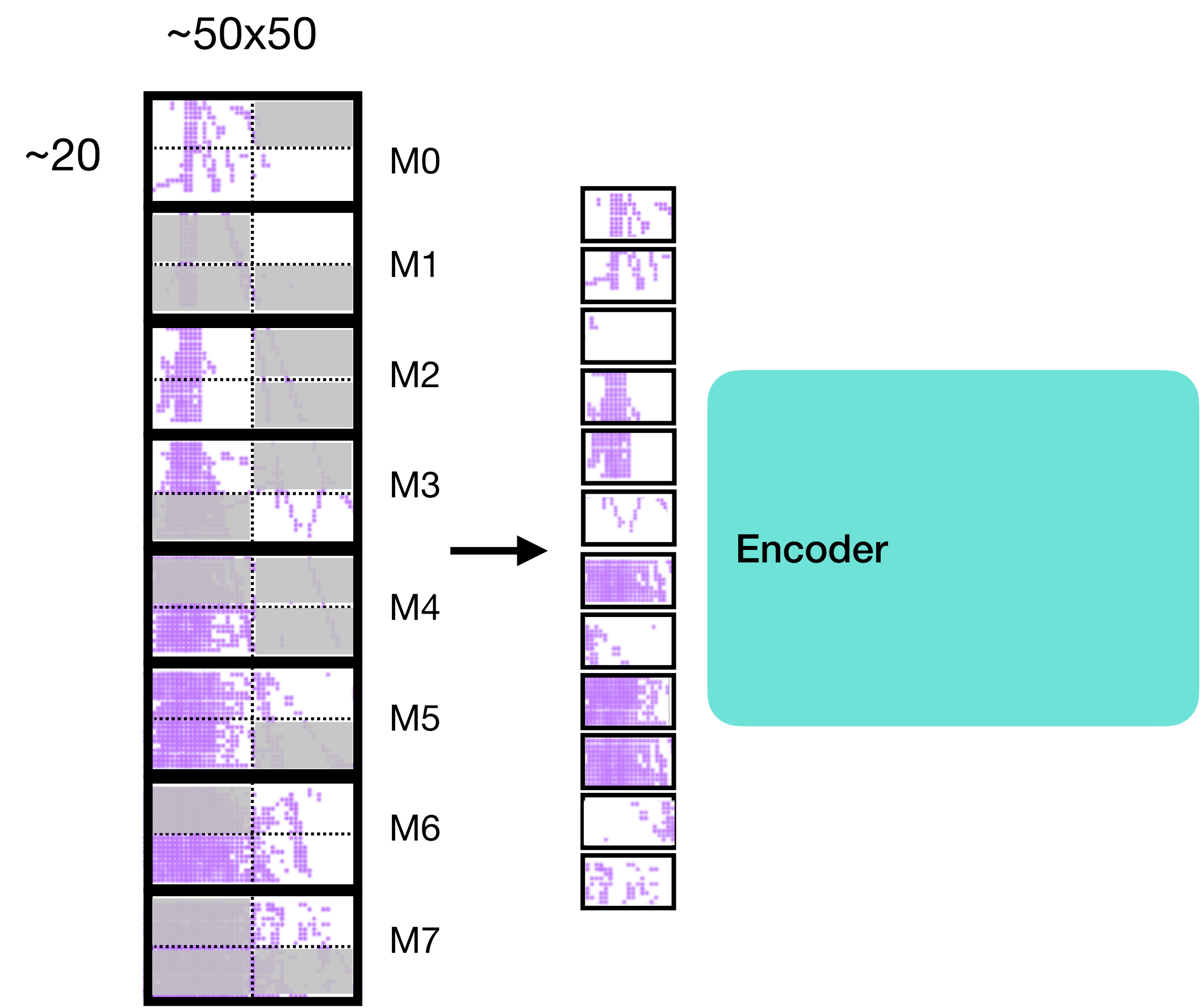
2



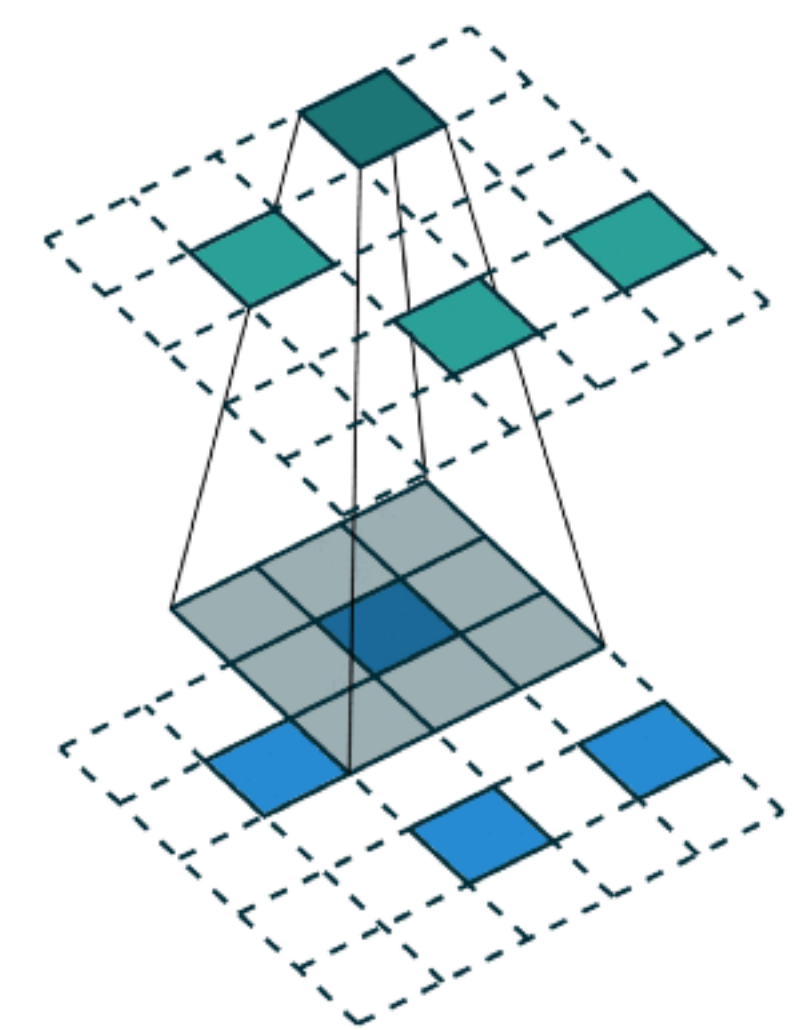


Pre-Training

Sparse Submanifold Neural Network

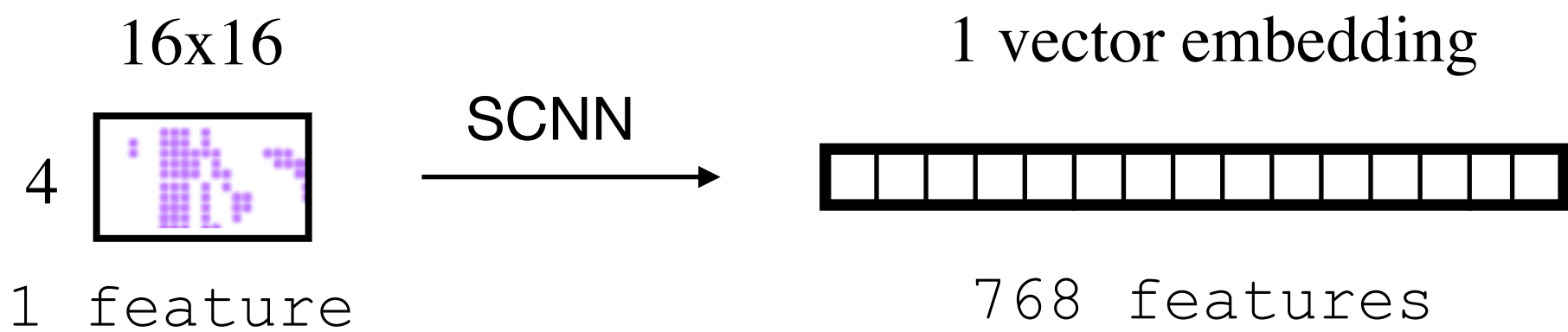


SCNN: Creates a vector embedding for every patch in the input



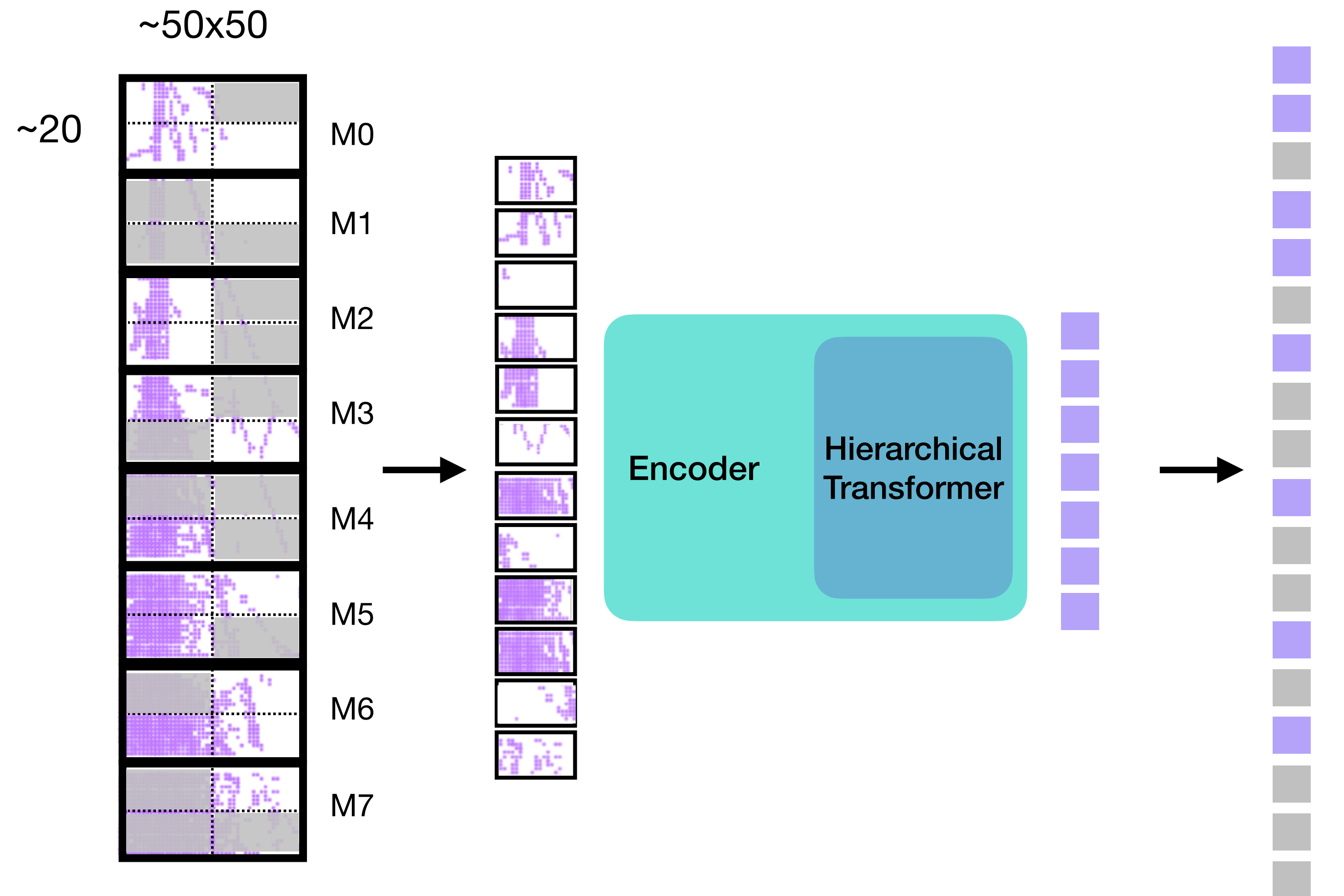
SCNN reduces the spatial coordinates while increasing the feature dimensions.

It performs convolution only on active voxels, effectively ignoring the vast empty regions typical in detector data.

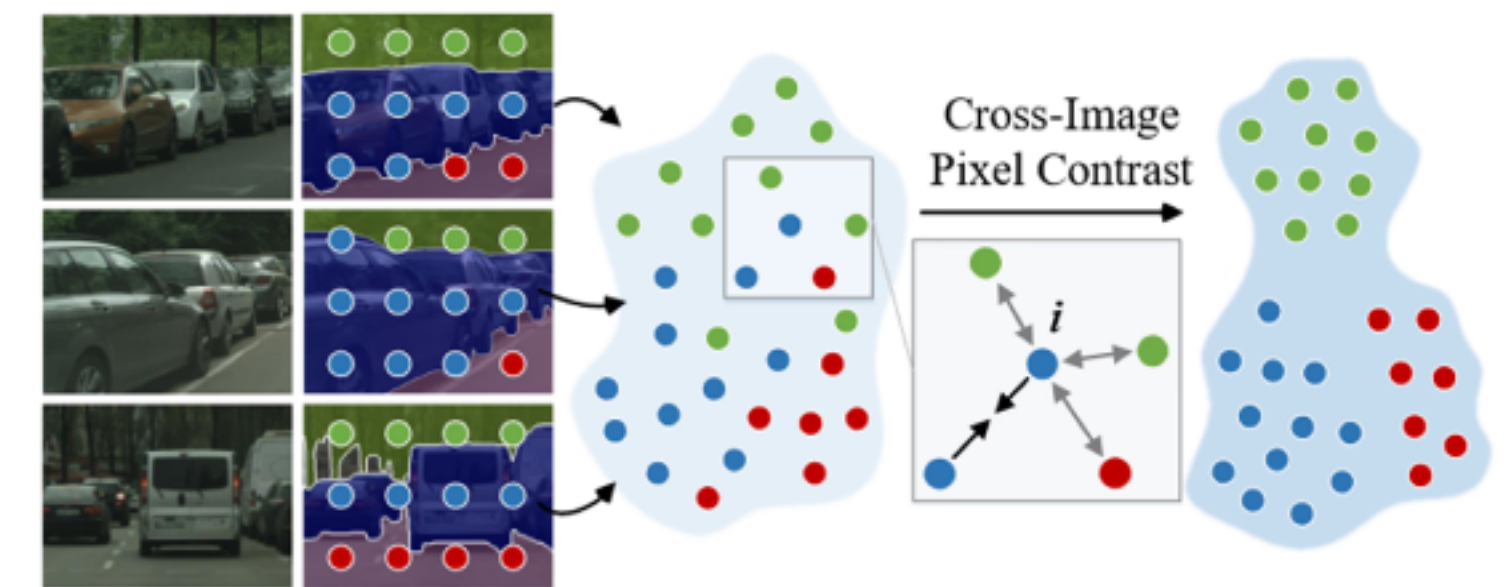


Pre-Training

Hierarchical Transformer



Contrastive Learning



Pixel-wise contrastive learning:

Each pixel (embedding) i is pulled closer to pixels of the same *PDG class*, but at the same time pushed far from pixels from other *PDG classes*