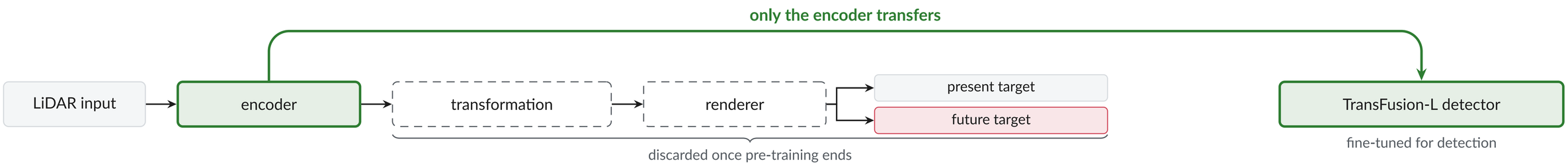
# What Does the Future Buy?

Separating projection from prediction in LiDAR pre-training

**Luca Brembilla**
Politecnico di Milano · luca.brembilla@polimi.it
ELLIS Summer School on Autonomous Driving

POLITECNICO MILANO 1863

Against full forecasting, present-only pre-training achieves competitive detection transfer on nuScenes and ONCE, higher NDS on nuScenes, and lower pre-training cost.




### THE TWO WE COMPARE

**Forecasting** — recurrent, ego-action-conditioned transformation
targets: present **+ future**

**Projection** — feed-forward, action-free projector
targets: present only

This pair changes *both* the transformation and the targets.

### THE TWO CONTROLS

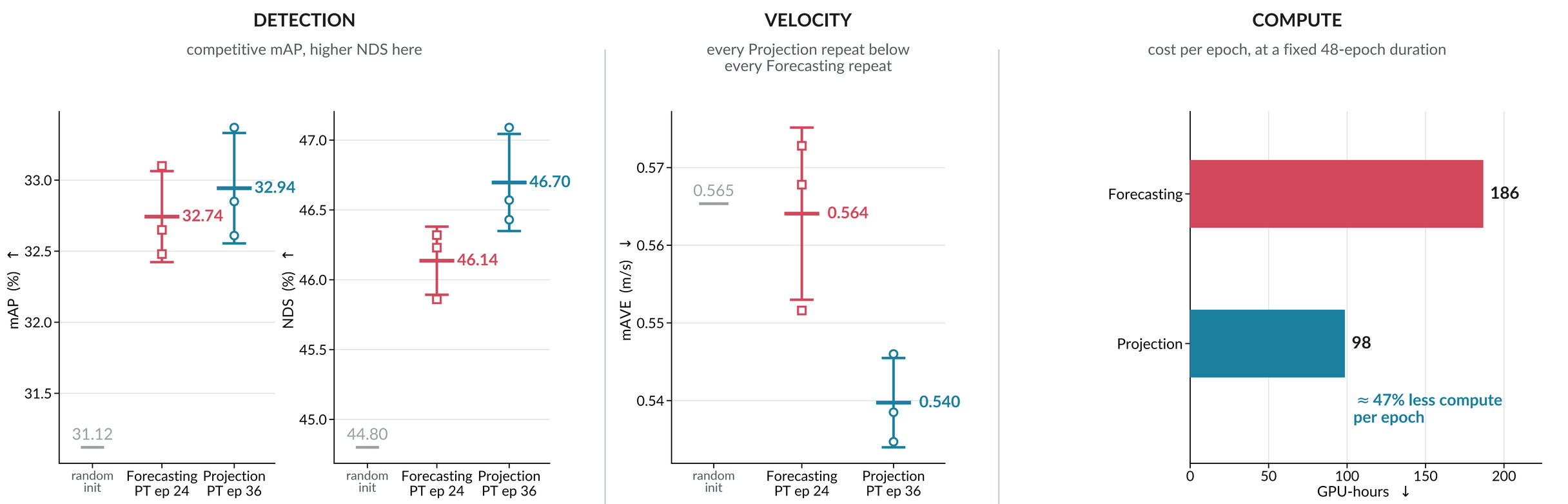
**Recurrent Projection** Forecasting with the future-target loss switched off
Isolates future supervision, architecture fixed

**Direct** no added transformation, present-only targets
Isolates the added transformation

**Testbed** TREND [1], future-predictive neural-field pre-training (NeurIPS 2025 Spotlight). **nuScenes:** a few-shot split, 176 labelled samples before balanced resampling. The encoder transfers to TransFusion-L at global batch 4 on a 16× schedule, evaluated at fine-tuning epoch 320. **ONCE:** the 20% subset with CenterPoint.

## Competitive detection, lower compute, and lower mAVE.

present-only Projection versus Forecasting · few-shot nuScenes · TransFusion-L




transfer at the two selected pre-training checkpoints, 3 fine-tuning repeats each: marker = one run, bar = mean, whiskers = ± sample SD

one fixed 48-epoch pre-training run

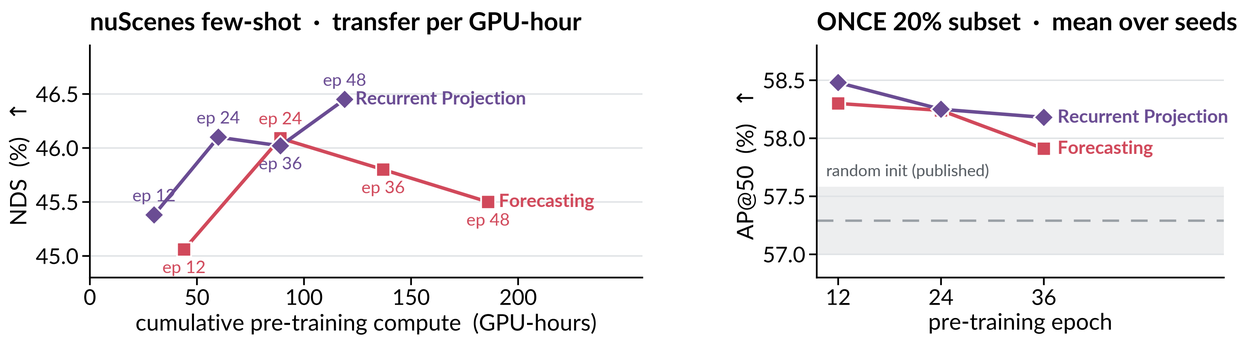
Checkpoints selected after examining detection results; one pre-training run per branch.

## What supports this simplification?

the two controls

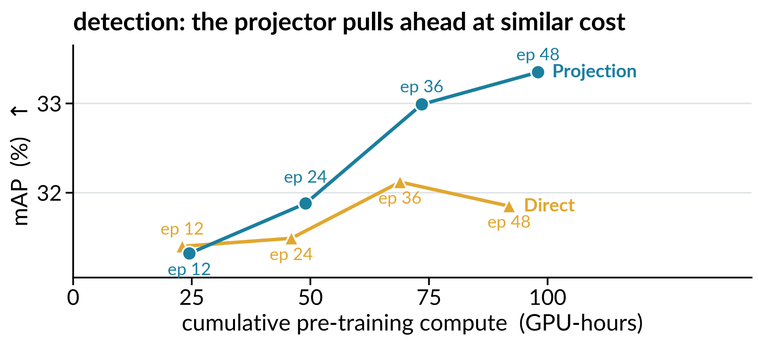
### FUTURE TARGETS

Recurrent Projection against Forecasting · future loss off against on, architecture fixed




### ADDED TRANSFORMATION

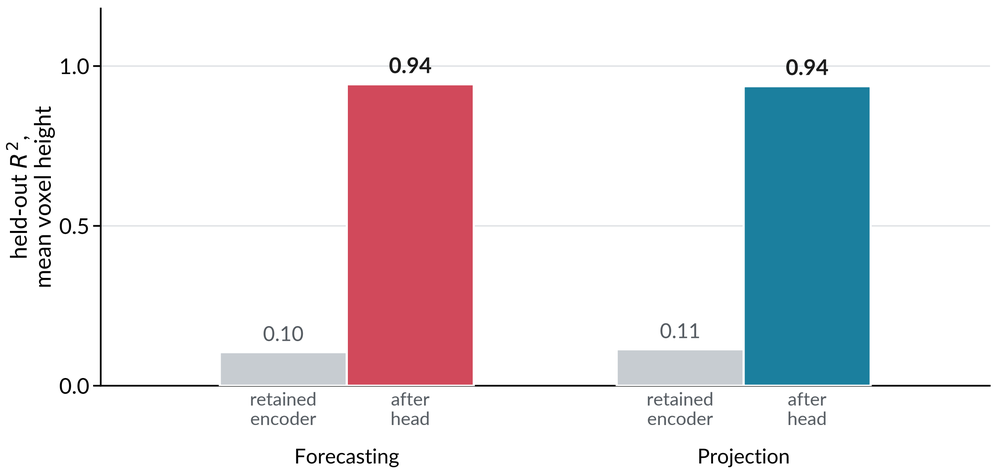
Projection against Direct · projector added, present-only targets




## What changes in the representation?

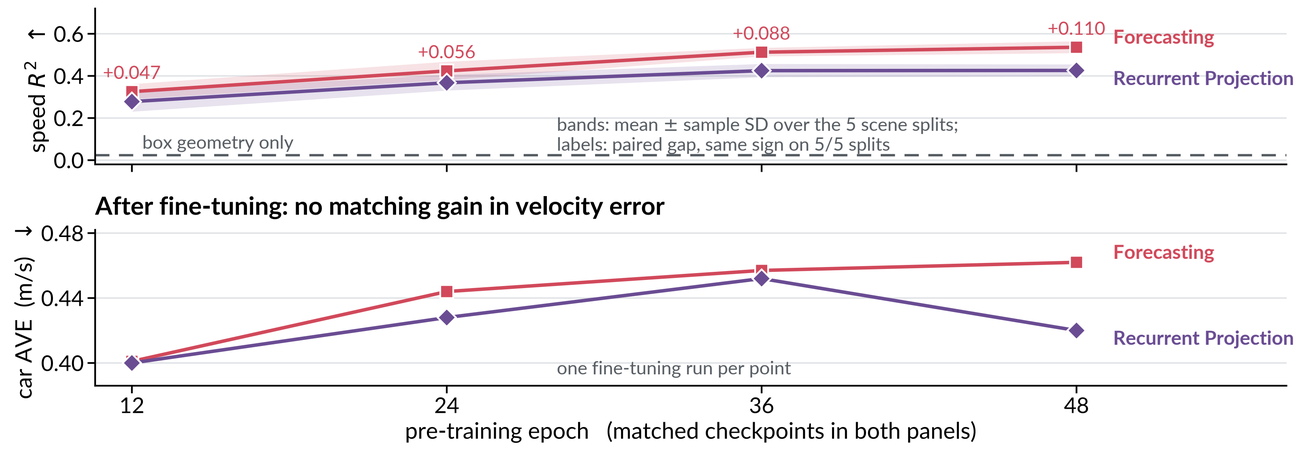
frozen representation probes and downstream velocity evaluation

**Mean voxel height becomes far more linearly accessible after the head**

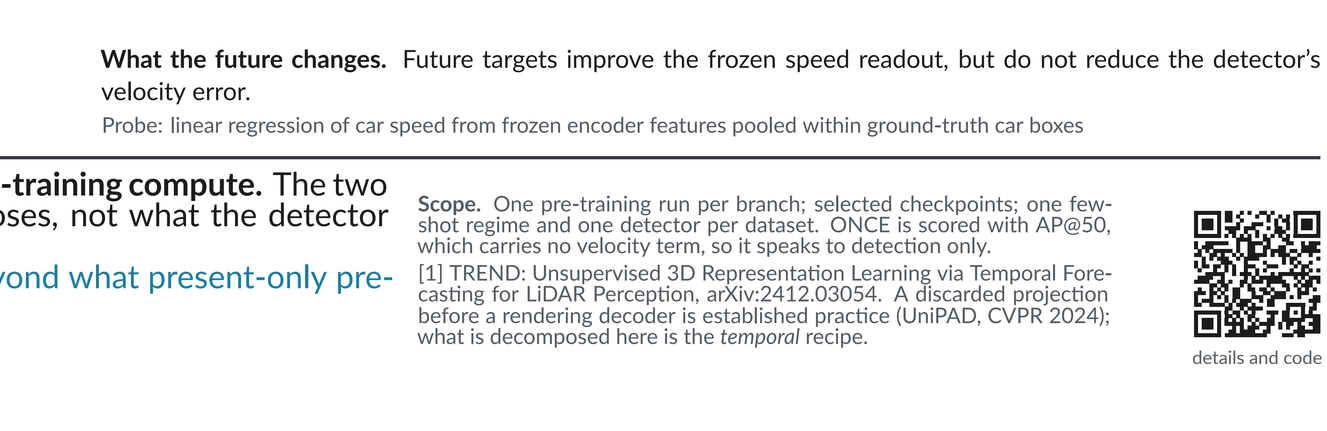



**What the head changes.** The head adapts the encoder's features to the rendering objective, turning them into a geometrically explicit code that fine-tuning then discards.

**Frozen encoder: speed is already accessible; future targets add more**




**After fine-tuning: no matching gain in velocity error**




**What the future changes.** Future targets improve the frozen speed readout, but do not reduce the detector's velocity error.

Probe: linear regression of car speed from frozen encoder features pooled within ground-truth car boxes

**A present-only projector supports competitive detection transfer with less pre-training compute.** The two ingredients come apart: future targets change what the frozen encoder exposes, not what the detector achieves.

Open question: which driving tasks benefit from future-target supervision beyond what present-only pre-training provides?

**Scope.** One pre-training run per branch; selected checkpoints; one few-shot regime and one detector per dataset. ONCE is scored with AP@50, which carries no velocity term, so it speaks to detection only.

[1] TREND: Unsupervised 3D Representation Learning via Temporal Forecasting for LiDAR Perception, arXiv:2412.03054. A discarded projection before a rendering decoder is established practice (UniPAD, CVPR 2024); what is decomposed here is the *temporal* recipe.

details and code