

BabyLand-72

Synthetic-to-Real Multi-View Infant Facial Landmark Estimation

Can infant-specific synthetic data train a landmark detector that transfers to real newborn images?

José Carlos Reyes-Hernández¹ · Antonia Alomar¹ · Ricardo Rubio² · Gemma Piella¹ · Federico Sukno¹

¹Universitat Pompeu Fabra · ²Hospital del Mar · Barcelona, Spain | josecarlos.reyes@upf.edu

ECCV 2026 · CDEL Workshop



Universitat
Pompeu Fabra
Barcelona



WE INTRODUCE

BabyLand-72 model

SynBaby-72 synthetic dataset

BabyLand-72 real dataset

Why infant landmarking is different

Different morphology

Infant craniofacial proportions differ from adult faces.

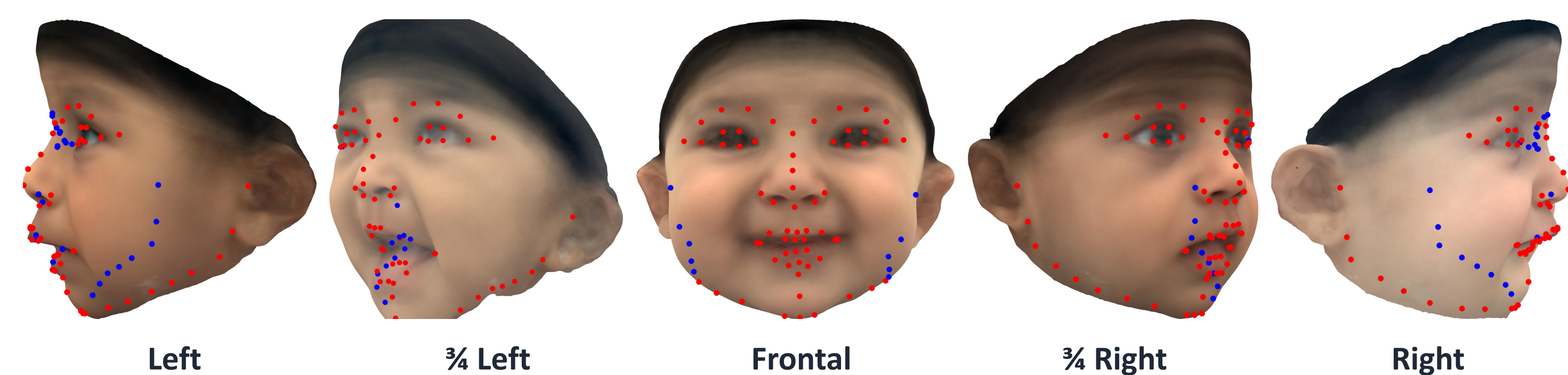
Limited annotated data

Real infant images are hard to collect and annotate at scale.

Pose + self-occlusion

Large yaw and self-occlusion make 2D contour landmarks ambiguous.

SynBaby-72 synthetic dataset



1,250 identities	16,250 images	13 yaw views	72 3D vertices
---------------------	------------------	-----------------	-------------------

Identity-disjoint train / validation / test splits

TRAIN 70%	VAL 15%	TEST 15%
875 identities · 11,375 images	187 · 2,431	188 · 2,444

BabyLand-72 dataset

Newborn photographs annotated with the same 72-landmark protocol.

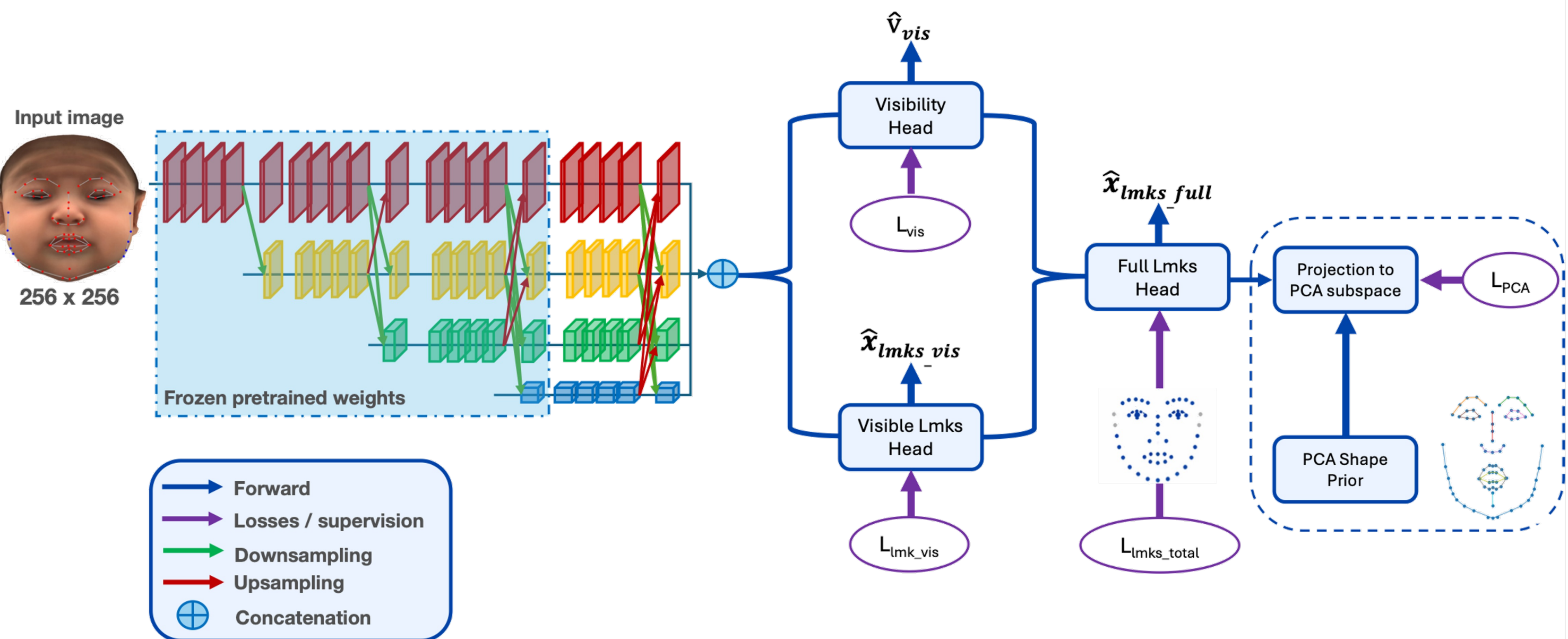


311 original images · 622 after mirroring · 5 pose groups

72 landmarks + visibility annotations

Five yaw-based orientation groups · coordinates stored as NaN when fully self-occluded

Visibility-aware BabyLand-72 model



HRNet-W18 backbone → visibility branch → visible-landmark branch → full 72-landmark branch → PCA shape regularization

No real infant images for training or fine-tuning

Synthetic → real evaluation

MODEL DEV.

SynBaby-72

train + validation
synthetic test evaluation

REAL DEV.

BabyLand-72

311 original images
no gradient updates

EXTERNAL TEST

InfAnFace

independent benchmark
no adaptation

Same 68 shared landmarks · P2L NME · Macro NME balances five pose groups

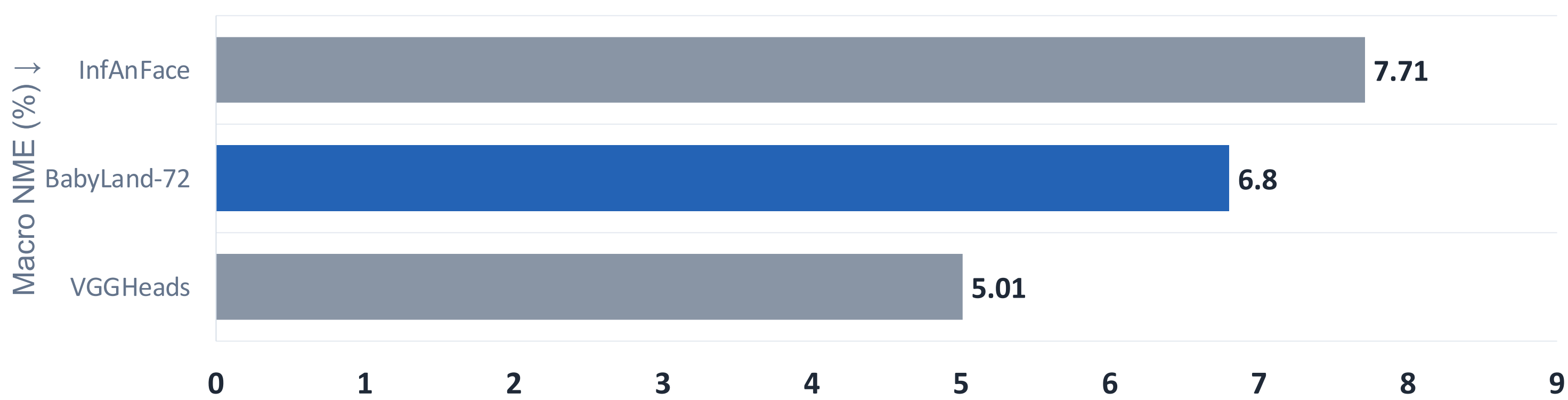
Qualitative comparison across pose



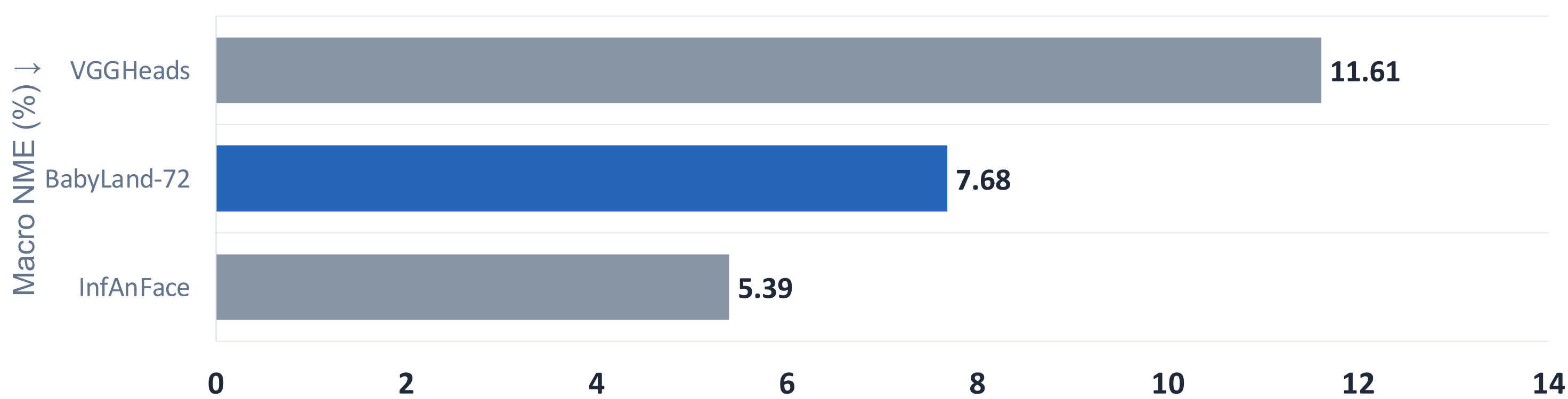
Synthetic-only training remains competitive

Macro NME gives equal weight to each pose category. Lower is better.

BabyLand-72 real dataset



InfAnFace external benchmark



100%

detection coverage
on both real datasets

20.43% / 0.70%

lowest Mean HD on BabyLand-72
best FR >20% on InfAnFace

What the evidence supports

● Synthetic-to-real transfer

Infant-specific synthetic supervision transfers without natural-image fine-tuning.

● Full detection coverage

Valid predictions for every evaluated image in both real datasets.

● Competitive, not universally best

Second-best Macro NME on both real datasets; the strongest baselines remain ahead.

TAKE-HOME

Infant-specific 3D synthetic supervision transfers competitively to real newborn images without real-infant fine-tuning.

FUTURE WORK

Broader synthetic diversity and realism · larger, more heterogeneous infant cohorts · targeted low-shot adaptation to real data