

Supplemental Material: Temporal Meta-Adaptor for Video Object Detection

Chi Wang^{1,2}
cwang38@qub.ac.uk

Yang Hua¹
y.hua@qub.ac.uk

Zheng Lu²
steven@anyvision.co

Jian Gao^{1,2}
jgao05@qub.ac.uk

Neil Robertson¹
n.robertson@qub.ac.uk

¹ EEECS/ECIT
 Queen's University Belfast
 Belfast, UK

² Anyvision
 Belfast, UK

1 Prediction Unit Design Details

A prediction unit (PU) generates video specific modulating parameters for internal features. The actual designs for predicting channel-wise shifting and scaling parameters β and γ are shown in Fig. 1.

w_β and w_γ are learnable weighting parameters for β and γ respectively. They are initialised to 0.

2 t-SNE visualization of video context

In Fig. 2, we show the full t-SNE visualization of video context.

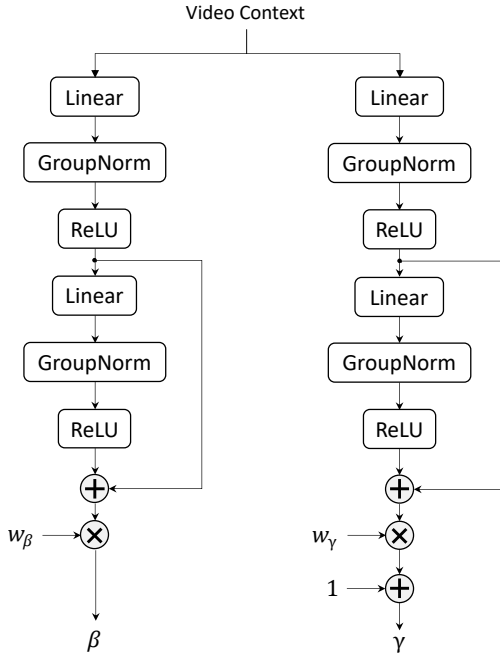


Figure 1: Prediction Unit for shifting parameter β and scaling parameter γ .

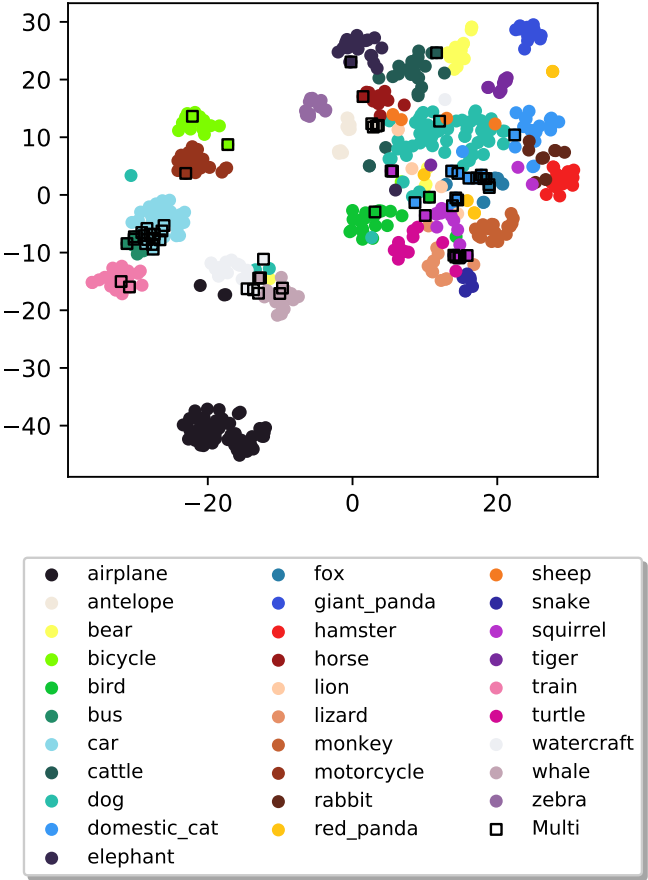


Figure 2: t-SNE visualisation of video context summarised from temporal features. Video context associated with multiple class labels is marked by a black box.