Rebuttal: Opening the Vocabulary of Egocentric Actions

Anonymous Author(s)
Affiliation
Address
email

Table 1: **CACNF** with S3D as backbone w/o and w/ OAP

Method	Top-1	Top-5
CACNF (S3D)	54.9	78.9
CACNF (S3D-OAP)	**56.2**	**81.7**

Table 2: **DETIC** vs **CLIP** vs proposed **AOP-Novel** (w/o temporal modeling)

Method	Base	Novel	HM
DETIC	12.0	2.3	3.9
CLIP	5.4	12.1	7.5
AOP-Novel	**19.6**	**18.4**	**22.7**

Table 3: Usefulness of $\mathcal{L}_{in}$ and object mixing

Loss	Temporal Gradients	Object Mixing	Top-1 (%)	Top-5 (%)
Supervised	✗	✗	49.3	77.3
$\mathcal{L}_{out}$	✗	✗	50.1	78.1
$\mathcal{L}_{out}$	✓	✓	52.1	79.3
$\mathcal{L}_{out} + \lambda\mathcal{L}_{in}$	✓	✓	**53.7**	**80.5**