AGES: Self-Attention and GCN Enhanced Audio-Visual Scene Recognition

(Search AGES on https://github.com website)

Supplementary Materials

Table 3 Performance of the single model on TAU

Categories		Audio	Visual	Audio and Visual (AGES)
Airport	Precision	11.7%	86.2%	98.9%
	Recall	43.3%	78.9%	99.3%
Bus	Precision	56.3%	56.8%	93.8%
	Recall	57.1%	66.4%	79.1%
Metro	Precision	22.0%	72.3%	99.5%
	Recall	50.5%	83.7%	99.1%
Square	Precision	52.7%	78.9%	98.8%
	Recall	76.7%	93.5%	99.2%
Metro Station	Precision	86.0%	92.1%	97.4%
	Recall	88.3%	99.2%	99.6%
Pedestrian	Precision	51.8%	62.4%	77.3%
	Recall	53.1%	89.0%	96.8%
Park	Precision	70.7%	90.2%	99.5%
	Recall	46.2%	70.0%	98.0%
Shopping Mall	Precision	61.1%	72.3%	98.9%
	Recall	48.7%	69.5%	81.6%
Street	Precision	85.8%	94.3%	99.1%
	Recall	79.7%	79.5%	99.5%
Tram	Precision	51.6%	64.1%	69.5%
	Recall	31.4%	50.5%	89.6%
Average	Precision	55.0%	74.0%	93.1%
	Recall	57.5%	78.0%	94.0%
	F1-Score	40.0%	77.8%	93.6%

Table 4 Evaluation of the three late fusion methods

Categories	Prediction Accuracy		
	Mean Value	Weighted Average	Exhaustive Grid Search
Airport	83.9%	84.1%	83.6%
Bus	64.7%	64.8%	66.7%
Metro	74.9%	75.1%	75.8%
Square	78.8%	78.8%	78.3%
Metro Station	91.2%	91.4%	91.7%
Pedestrian	83.8%	84.0%	83.9%
Park	98.2%	98.2%	98.3%
Shopping Mall	83.7%	83.8%	84.0%
Street	95.7%	95.8%	95.7%
Tram	62.1%	62.0%	63.2%
Average	82.2%	82.3%	82.6%

Table 5 Evaluation of feature fusion methods

Categories		Fusion Method		
		Early Fusion	Fusion with GCN	AGES
Airport	Precision	95.1%	98.9%	98.9%
	Recall	98.3%	89.4%	99.3%
Bus	Precision	81.6%	93.2%	93.8%
	Recall	84.1%	75.5%	79.1%
Metro	Precision	89.8%	86.6%	99.5%
	Recall	99.2%	95.2%	99.1%
Square	Precision	98.8%	87.5%	98.8%
	Recall	96.0%	96.1%	99.2%
Metro Station	Precision	99.6%	94.4%	97.4%
	Recall	95.7%	96.4%	99.6%
Pedestrian	Precision	91.6%	70.4%	77.3%
	Recall	79.9%	87.7%	96.8%
Park	Precision	99.1%	93.8%	99.5%
	Recall	98.7%	88.3%	98.0%
Shopping Mall	Precision	82.5%	92.0%	98.9%
	Recall	94.9%	77.5%	81.6%
Street	Precision	99.5%	96.3%	99.1%
	Recall	98.3%	93.3%	99.5%
Tram	Precision	80.2%	72.6%	69.5%
	Recall	70.1%	86.7%	89.6%
Average	Precision	91.7%	92.3%	93.1%
	Recall	91.5%	88.6%	94.0%
	F1-Score	91.6%	90.4%	93.6%