

TABLE II: Within-dataset evaluation of various models using four MultiMatch metrics and D-score on the OSIE test images. The top three results in order are marked in red, green and blue, respectively. H represents the Inter-observer D-score and the D-scores within $H \pm 0.25$ are shown in bold.

Method	MultiMatch Scores \uparrow				D-score
	Vector	Direction	Length	Position	within ($H \pm 0.25$)?
Inter-observer	0.8880	0.6916	0.8926	0.8322	0.5364(H)
Itti [10]	0.8827	0.6719	0.8753	0.7942	0.0286
SGC [28]	0.8726	0.6286	0.8518	0.7674	0.2744
CLE [27]	0.8704	0.5739	0.8682	0.8010	0.5338
Le Meur [12]	0.8882	0.6545	0.8897	0.7845	0.4843
STAR-FC [15]	0.8717	0.6126	0.8765	0.7874	0.2005
IOR-ROI [17]	0.8819	0.6969	0.8900	0.8047	0.4537
SQA [25]	0.8742	0.6444	0.8770	0.8258	0.2373
ScanFormer	0.8903	0.6584	0.8924	0.8122	0.5023
ScanFormer*	0.8926	0.6555	0.8936	0.8142	0.4057