

# LaScA: Language-Conditioned Scalable Modelling of Affective Dynamics

Kosmas Pinitas, Ilias Maglogiannis

## The Challenge

Deep embeddings lack **Interpretability** while handcrafted features lack **Contextual Understanding**. Expert knowledge is difficult to integrate without sacrificing **Performance**.

## Contributions

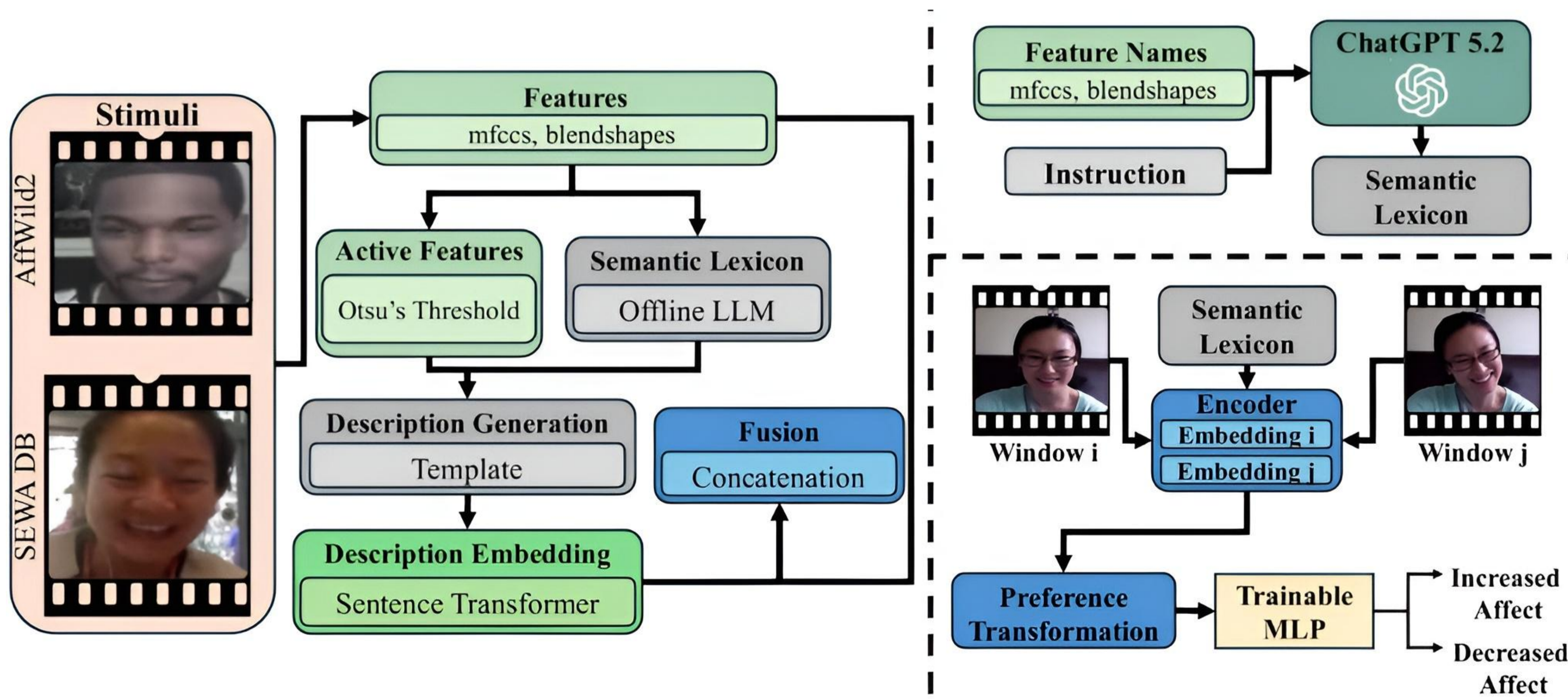
- Distribution-aware framework
- Higher-order descriptor
- Comparative with SOTA

## Results

Modality	Model	Arousal				Valence			
		3s		5s		3s		5s	
		10%	20%	10%	20%	10%	20%	10%	20%
Visual	VGGFace2	0.64	0.63	0.66	0.71	0.57	0.60	0.65	0.72
	SwinFace	<b>0.67</b>	<b>0.67</b>	<u>0.70</u>	<b>0.74</b>	<b>0.61</b>	<b>0.63</b>	<b>0.69</b>	<u>0.73</u>
	MAE-Face	0.65	0.64	<b>0.71</b>	0.72	0.58	<u>0.61</u>	<u>0.67</u>	0.71
	LaScA (ours)	<u>0.66</u>	0.65	<u>0.69</u>	<b>0.74</b>	0.59	<u>0.62</u>	<u>0.68</u>	<b>0.74</b>
Audio	Wav2Vec2	<b>0.64</b>	<u>0.66</u>	<u>0.62</u>	<u>0.71</u>	<b>0.55</b>	<b>0.57</b>	<b>0.58</b>	<b>0.60</b>
	MAE-Audio	0.61	0.65	0.61	0.69	0.53	0.55	<u>0.57</u>	<b>0.60</b>
	LaScA (ours)	0.62	<b>0.67</b>	<b>0.63</b>	<b>0.72</b>	<u>0.54</u>	<u>0.56</u>	<b>0.58</b>	0.58
Multimodal	HiCMAE	<u>0.64</u>	<u>0.68</u>	<b>0.70</b>	<b>0.75</b>	<b>0.60</b>	<b>0.61</b>	<b>0.61</b>	<b>0.63</b>
	MMA-DFER	0.62	<b>0.69</b>	0.68	<b>0.75</b>	<b>0.60</b>	<u>0.60</u>	<u>0.60</u>	<b>0.63</b>
	LaScA (ours)	<b>0.65</b>	<b>0.69</b>	0.67	<u>0.74</u>	<u>0.58</u>	<u>0.60</u>	<b>0.61</b>	0.61

Modality	Model	Arousal				Valence			
		3s		5s		3s		5s	
		10%	20%	10%	20%	10%	20%	10%	20%
Multimodal	LaScA								
	MPNet(F)	0.63	0.67	0.65	<b>0.74</b>	0.56	0.59	0.60	<b>0.61</b>
MPNet(F)	Feature	0.63	0.67	0.65	<b>0.74</b>	0.56	0.59	0.60	<b>0.61</b>
	LLM (ours)	<b>0.65</b>	<b>0.69</b>	<b>0.67</b>	<b>0.74</b>	<b>0.58</b>	<b>0.60</b>	<b>0.61</b>	<b>0.61</b>
DRoBERTa(F)	Feature	0.64	0.68	0.66	<b>0.74</b>	0.56	0.59	0.60	0.61
	LLM (ours)	<b>0.65</b>	<b>0.70</b>	<b>0.67</b>	<b>0.74</b>	<b>0.59</b>	<b>0.60</b>	<b>0.61</b>	<b>0.63</b>
MiniLM(F)	Feature	0.64	0.67	0.65	<b>0.72</b>	0.56	<b>0.57</b>	0.60	0.60
	LLM (ours)	<b>0.65</b>	<b>0.69</b>	<b>0.67</b>	<b>0.72</b>	<b>0.58</b>	<b>0.57</b>	<b>0.61</b>	<b>0.61</b>

## The LaScA Framework



## Conclusions

- Semantic Lexicon improved performance by **1-2%**
- LaScA was competitive with SOTA
- Training was minimal (**130k params**)

