UNIVERSITY OF CAMBRIDGE
Department of Computer Science and Technology

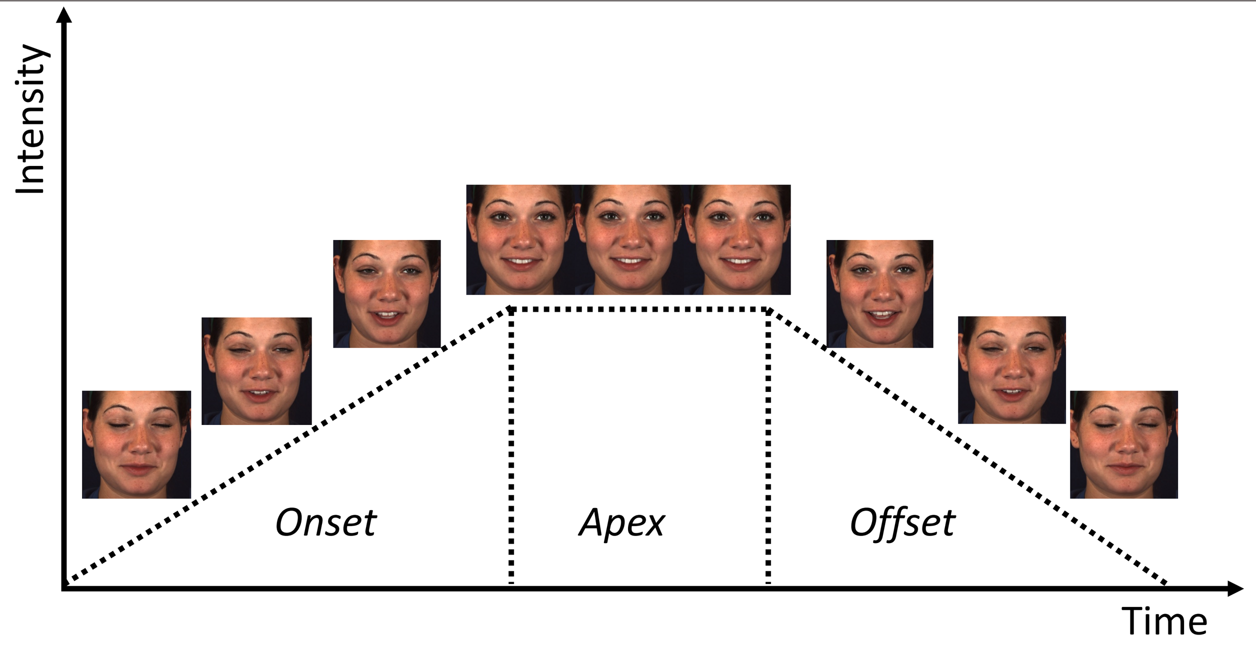
ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

AULA-Caps: Lifecycle-Aware Capsule Networks for Spatio-Temporal Analysis of Facial Actions

Nikhil Churamani[1], Sinan Kalkan[2] and Hatice Gunes[1]
[1]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
[2]Department of Computer Engineering, Middle East Technical University, Ankara, Turkey

Scan to Contact Us

Motivation



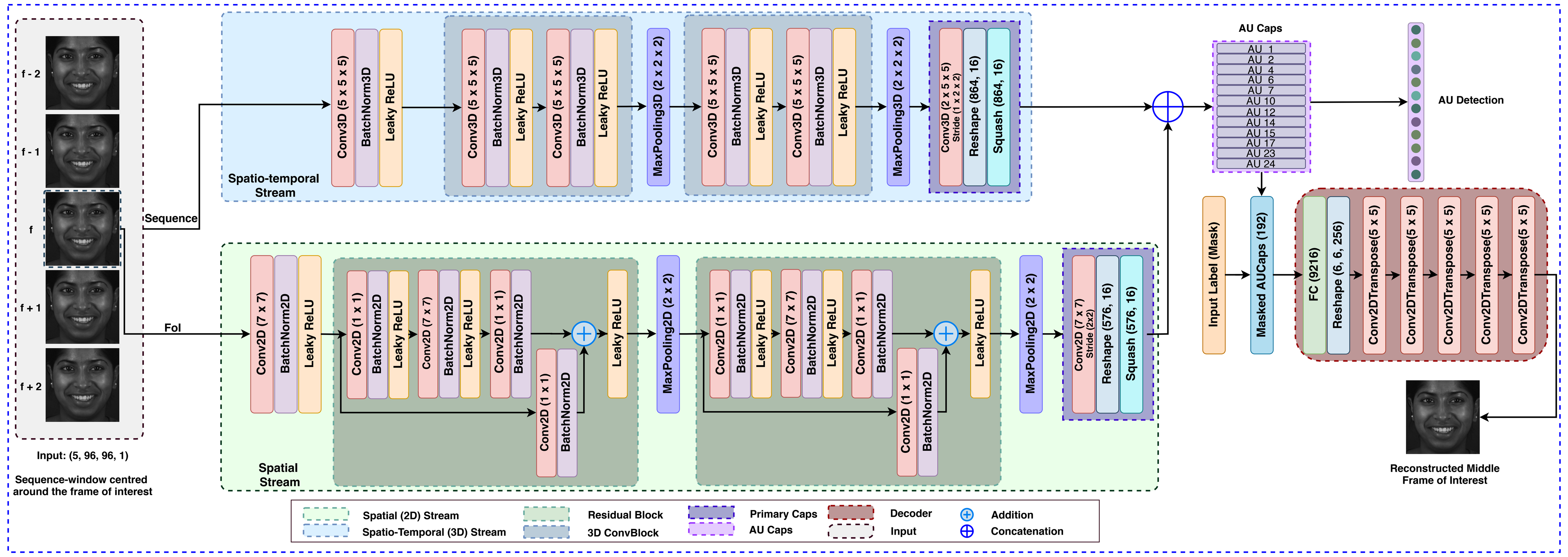
- Facial Action Unit (AU) Activation follows a temporal evolution: the ***AU Lifecycle***.
  - Facial muscles contract to form the ***onset*** phase.
  - Complete contraction at the ***apex*** state.
  - Muscles start to relax in the ***offset*** phase.

Lifecycle Awareness

- Spatial and Spatio-temporal capsules **weighed competitively** based on their contribution.
- Precedence given to **spatial feature primitives** during **apex** segments.
- **Spatio-temporal feature primitives** capture differences in **contiguous frames** for **off-peak** segments.
- **Ablations** highlight **importance** of **combining** spatial and spatio-temporal features.

Model	Avg. F1-Score
2D Stream AULA-Caps	0.580
3D Stream AULA-Caps	0.550
AULA-Caps (N=2)	**0.645**

Proposed Approach



Results on BP4D and GFT Datasets

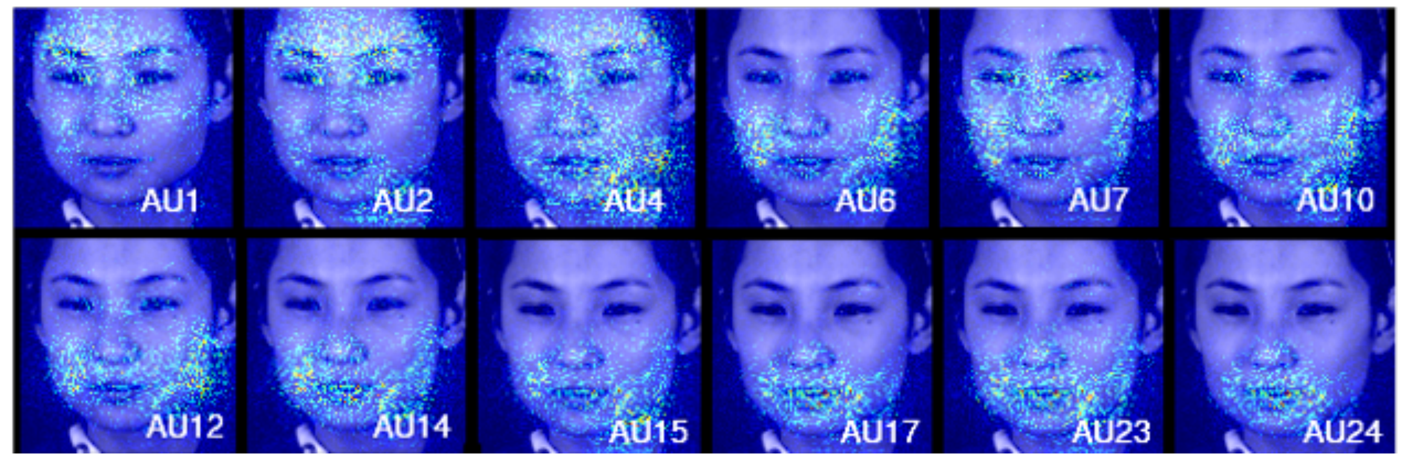
TABLE I: Performance Evaluation (F1-Scores) on BP4D. **Bold** values denote best while [*bracketed*] denote second-best values for each row.

AU	CNN-LSTM [6]	EAC [7]	ROI [33]	CapsNet [24]	JÂA [34]	SRERL [17]	STRAL [9]	AULA-Caps [*Ours*]
1	0.314	0.390	0.362	0.468	[*0.538*]	0.469	0.482	**0.562**
2	0.311	0.352	0.316	0.291	**0.478**	0.453	[*0.477*]	0.465
4	**0.714**	0.486	0.434	0.529	[*0.582*]	0.556	0.581	0.573
6	0.633	0.761	0.771	0.753	[*0.785*]	0.771	0.758	**0.796**
7	0.771	0.729	0.737	0.776	0.758	**0.784**	[*0.781*]	0.765
10	0.450	0.819	**0.850**	0.824	0.827	0.835	0.816	[*0.843*]
12	0.826	0.862	0.870	0.850	**0.882**	[*0.876*]	[*0.876*]	0.874
14	**0.729**	0.588	0.626	0.657	0.637	0.639	0.605	[*0.718*]
15	0.340	0.375	0.457	0.337	0.433	**0.522**	[*0.502*]	0.457
17	0.539	0.591	0.580	0.606	0.618	0.639	[*0.640*]	**0.694**
23	0.386	0.359	0.383	0.369	0.456	0.471	**0.512**	[*0.495*]
24	0.370	0.358	0.374	0.431	0.499	[*0.533*]	**0.552**	0.502
Avg.	0.532	0.559	0.564	0.574	0.624	0.629	[*0.632*]	**0.645**

TABLE II: Performance Evaluation (F1-Scores) on GFT. **Bold** values denote best while [*bracketed*] denote second-best values for each row. *Averaged for 10 AUs.

AU	CRD [23]	ANet [6]	JÂA [34]	CNN-LSTM [6]	AULA-Caps [*Ours*]
1	[*0.437*]	0.312	**0.465**	0.299	0.313
2	0.449	0.292	[*0.493*]	0.257	**0.498**
4	0.198	**0.719**	0.192	[*0.689*]	0.297
6	0.746	0.645	**0.790**	0.673	[*0.775*]
7	0.721	0.671	–	[*0.725*]	**0.772**
10	**0.765**	0.426	[*0.75*]	0.670	0.749
12	[*0.798*]	0.731	**0.848**	0.751	0.785
14	0.500	[*0.691*]	0.441	**0.807**	0.236
15	0.339	0.279	0.335	**0.435**	[*0.371*]
17	0.170	[*0.504*]	–	0.491	**0.592**
23	0.168	0.348	**0.549**	0.350	[*0.522*]
24	0.129	0.390	[*0.507*]	0.319	**0.530**
Avg.	0.452	0.500	0.537*	**0.539**	[*0.537*]

Visualisations



- ***Saliency Maps*** depicting pixels in the FoIs that contribute the most to model predictions.
- Additional activity witnessed for **AU 4** in the lower face region due to co-activation with **AU 7** and **AU 17.**
- Understanding **co-occurrence patterns** can help **improve model performance.**

Conclusion and Future Work

- Combining **spatial** and **spatio-temporal** feature primitives improves model performance.
- **Selectively focusing** on spatial and spatio-temporal features through capsule routing enables robustness.
- Model performance **sensitive** to sequence window length.
- Dynamically **adapting** window-size based on specific AU lifecycles using **anchor frames** (Lu *et al.* 2020).
- Data Imbalance major hurdle for multi-label classification problems.