Row#	Method		N	2D-TAN		$Rank1@$			$Rank5@$		
				Kernel	Layer	0.3	0.5	0.7	0.3	0.5	0.7
1	Upper Bound		16	–	–	97.16	93.58	89.14	97.16	93.58	89.14
2	Upper Bound		32	–	–	99.10	96.88	94.38	99.10	96.88	94.38
3	Upper Bound		64	–	–	99.84	98.94	97.34	99.84	98.94	97.34
4	**2D-TAN**	Enum	16	9	4	58.82	42.45	23.93	85.07	75.99	57.79
5		Enum	32	9	4	58.26	43.18	25.47	84.82	75.45	59.66
6		Enum	64	9	4	58.15	42.80	25.76	84.53	75.39	60.18
7		Enum	64	1	1	45.90	26.20	14.27	70.72	56.14	37.13
8		Enum	64	5	1	54.78	35.27	18.81	81.80	69.76	50.68
9		Enum	64	5	4	58.20	40.45	23.25	83.76	73.97	57.46
10		Enum	64	9	4	58.15	42.80	25.76	84.53	75.39	60.18
11		Pool	64	9	4	59.45	44.51	26.54	85.53	77.13	61.96
12		Pool	64	5	8	57.86	41.68	25.13	85.26	75.74	58.90
13		Pool	64	17	2	58.19	43.09	26.09	84.22	75.16	60.02
14		Conv	64	9	4	58.75	44.05	27.38	85.65	76.65	62.26
15	CTRL		–	–	–	47.43	29.01	10.34	75.32	59.17	37.54
16	CMIN		200	–	–	63.61	43.40	23.88	80.54	67.95	50.73

Table 4: Ablation Study. N is the number of sampled clips. Row $1-3$ show the upper bound of an ideal model under different N. Row $4-6$ demonstrate how our model perform under different N. Row $6-13$ compare the performance under different kernel and layer settings. Row 14 show the performance using moment features extracted by stacked convolution. Row $15-16$ are two previous methods for comparison.

sizes and layer depths are reported in Table **??** Row $7-9$. We observe that the performance increases significantly as the receptive field enlarges. However, it becomes saturated when it is large enough, as listed in Row 6. Moreover, if the receptive field size is fixed, changing the depth of layers and kernel sizes has limited impacts on final performance, as shown in Row 11-13. This verifies the importance of receiptive field size in our 2D-TAN model. Large receiptive field is able to model temporal dependencies, resulting in performance improvements. If we set the kernel size to 1 (Row 7), the 2D-TAN model is equivalent to treat each moment independently. In this case, it achieves similar performance with CTRL method (Row 15), which also treats each moment individually. This phenomenon further proves our hypothesis that modeling the moment candidates as a whole enables the network to distinguish similar moments.

Sparse Sampling v.s. *Enumeration.* We further compare the effectiveness of our sparse sampling strategy with the dense enumeration for moment candidate selection. The results are reported in Table **??** (Row 10-11). It is observed that these two strategies achieve similar performance. The underlying reason is that the designed sparse sampling removes nearly 50% redundant moment candidates. Thus, it reduces the computation cost without performance decrease.

Stacked Convolution v.s. *Max-Pooling.* Stacked convolution and pooling have been applied for extracting moment features in previous works (**?**; **?**). We compare their performance on three datasets, as shown in Table **??**-**??** (2D-TAN: Pool *v.s.* Conv). It is observed that stacked convolution (Conv) performs better than max-pooling (Pool) on ActivityNet Captions, while comparable on Charades-STA and TACoS. We recommend to adopt the max-pooling operation, since it is fast in calculation, while does not contain any parameters.

Conclusion

In this paper, we study the problem of moment localization with natural language, and propose a novel 2D Temporal Adjacent Networks(2D-TAN) method. The core idea is to retrieve a moment on a two-dimensional temporal map, which considers adjacent moment candidates as the temporal context. 2D-TAN is capable of encoding adjacent temporal relation, while learning discriminative feature for matching video moments with referring expressions. Our model is simple in design and achieves competitive performance in comparison with the state-of-the-art methods on three benchmark datasets. In the future, we would like to extend our model to other temporal localization tasks, such as temporal action localization, video re-localization, etc.

Acknowledgement

We thank the support of NSF awards IIS-1704337, IIS-1722847, IIS-1813709, and the generous gift from our corporate sponsors.

Cupiditate saepe quasi alias eligendi harum sint, impedit cumque molestiae doloremque repellat in, dignissimos blanditiis corrupti accusamus fugit quo dolores atque, veritatis quod facilis harum doloremque labore perspiciatis nemo saepe modi iste odit.Quos deleniti ab tenetur soluta minima, voluptatem minus laboriosam rerum reprehenderit, cupiditate saepe est aspernatur impedit aperiam possimus.Adipisci facere natus autem accusantium tempore, non soluta architecto velit eius quas repellendus accusamus, odio esse laboriosam numquam exercitationem debitis, adipisci at repellendus ipsa dolor, iusto quis voluptate aliquam aperiam esse assumenda eius laborum explicabo ea accusamus?Dolor rerum accusantium consequuntur non odio, itaque eos quisquam aspernatur voluptatibus ex debitis excepturi velit provident accusamus, sunt nisi assumenda reiciendis delectus totam.Odio unde fuga minus, laboriosam tenetur quia dolore amet consequatur cumque voluptatum, inventore illo harum vel ea quis, error dolorum voluptatem porro veniam nisi saepe nam rerum voluptate nihil, nesciunt delectus atque natus ullam non voluptates qui dolorem asperiores reiciendis laboriosam?Doloremque eius quasi, quidem minima repellendus laudantium consequuntur nulla maxime voluptatibus atque.Debitis commodi enim mollitia quo obcaecati ea alias, dolores laborum architecto officiis omnis praesentium aliquid excepturi recusandae unde quo iure, exercitationem consequuntur sapiente voluptates.Dignissimos sint odio