Rubric-Constrained Figure Skating Scoring Supplementary

This supplemental section contains all rubrics utilized in our experiments, including the full set of hand-negated rubric items used to enhance visual-text pretraining (Sec. 1). It also presents additional experiments, such as ablation studies on the margin for each pretraining method used in our experiments and the effect of visual-text pretraining on element-rubric item alignment compared to ground truth quality (Sec. 2.1), sample cross-attention maps showing the impact of implicit segmentation regularization (Sec. 2.2), visualization of what our element queries attend to (Sec. 2.3), and results on an additional dataset, FS1000 (Sec. 2.4).

1. Figure Skating Technical Element Scoring Rubric Variants

1.1. Simplified Rubric

The authors simplify the original rubric contents shown in Fig. 1 and Fig. 2 to be more understandable to a non-expert and text encoder.

List of simplified rubric items (positive):

1. "jump that is unexpected or creative or difficult entry"
2. "jump with clear recognizable steps/free skating movements immediately preceding element"
3. "jump with varied position in the air / delay in rotation"
4. "jump with good height and distance"
5. "jump with good extension on landing or creative exit"
6. "jump with good flow from entry to exit including jump combinations / sequences"
7. "jump that is effortless throughout"
8. "jump matched to the musical structure"
9. "spin with good speed or acceleration during spin"
10. "centering a spin quickly"
11. "spin with balanced rotations in all positions"
12. "spin with clearly more than required number of revolutions"
13. "spin good, strong position(s) and height and air/landing position in flying spins"
14. "creative and original spin"
15. "good control throughout all phases of spin"
16. "spin matched to the musical structure"
17. "step sequence with good energy and execution"
18. "good speed or acceleration during step sequence"
19. "well executed various steps during the step sequence"
20. "step sequence with deep clean edges including entry and exit of all turns"
21. "good control and commitment of the whole body maintaining accuracy of steps in step sequence"
22. "creative and original step sequence"
23. "effortless throughout step sequence"
24. "step sequence enhances the musical structure"
25. "choreographic step sequence with good flow, energy and execution"
26. "good speed or acceleration during choreographic step sequence"
27. "choreographic step sequence with good clarity and precision"
28. "choreographic step sequence with good control and commitment of whole body"
29. "creative and original choreographic step sequence"
30. "effortless throughout choreographic step sequence"
31. "choreographic step sequence reflecting concept or character of the program"
32. "choreographic step sequence enhancing the musical structure"

List of simplified rubric items (negative):

FOR + 1 : 2 bullets **FOR + 2 : 4 bullets** **FOR + 3 : 6 or more bullets**

Singles

Jump Elements	1) unexpected / creative / difficult entry 2) clear recognizable (**difficult** for jump preceded by steps/movements of the Short Program) steps/free skating movements immediately preceding element 3) varied position in the air / delay in rotation 4) good height and distance 5) good extension on landing / creative exit 6) good flow from entry to exit including jump combinations / sequences 7) effortless throughout 8) element matched to the musical structure
Spins	1) good speed or acceleration during spin 2) ability to center a spin quickly 3) balanced rotations in all positions 4) clearly more than required number of revolutions 5) good, strong position(s) (including height and air/landing position in flying spins) 6) creativity and originality 7) good control throughout all phases 8) element matched to the musical structure
Step Sequences	1) good energy and execution 2) good speed or acceleration during sequence 3) use of well executed various steps during the sequence 4) deep clean edges (including entry and exit of all turns) 5) good control and commitment of the whole body maintaining accuracy of steps 6) creativity and originality 7) effortless throughout 8) element enhances the musical structure
Choreographic Sequences	1) good flow, energy and execution 2) good speed or acceleration during sequence 3) good clarity and precision 4) good control and commitment of whole body 5) creativity and originality 6) effortless throughout 7) reflecting concept/character of the program 8) element enhances the musical structure

Figure 1. Original positive rubric items from 2016/2017 grade of execution scoring protocol that award points from [2]

1. “combo contained only one jump”
2. “downgraded jump”
3. “no required preceding steps/movements prior to jump”
4. “under-rotated jump”
5. “break between required steps/movements & jump/only 1 step/movement preceding jump”
6. “jump lacking rotation including half loop in a combo”
7. “fall during jump”
8. “jump has poor speed, height, distance, air position”
9. “landing on two feet in a jump”
10. “touch down with both hands in a jump”
11. “stepping out of landing in a jump”
12. “touch down with one hand or free foot in a jump”
13. “2 three turns in between in a jump combo”
14. “loss of flow/direction/rhythm between jumps in combo or sequence”
15. “severe wrong edge take-off in jump”
16. “jump with weak landing (bad pos./wrong edge/scratching etc)”
17. “jump with unclear or wrong edge take off”
18. “jump with poor take-off”
19. “jump with long preparation”
20. “flying spin where prescribed air position not attained”
21. “spin with poor/awkward, unaesthetic position(s)”
22. “fall during spin”

REDUCTIONS FOR ERRORS			
JUMP ELEMENTS			
SP: Combo of one jump **final GOE must be**	**-3**	Downgraded (sign <<)	**-2 to -3**
SP: No required preceding steps/movements	**-3**	Under-rotated (sign <)	**-1 to -2**
SP: Break between required steps/movements & jump/only **1** step/movement preceding jump	**-1 to -2**	Lacking rotation (no sign) including half loop in a combo	**-1**
Fall	**-3**	Poor speed, height, distance, air position	**-1 to -2**
Landing on two feet in a jump	**-3**	Touch down with both hands in a jump	**-2**
Stepping out of landing in a jump	**-2 to -3**	Touch down with one hand or free foot	**-1**
2 three turns in between (jump combo)	**-2**	Loss of flow/direction/rhythm between jumps (combo/seq.)	**-1 to -2**
Severe wrong edge take off F/Lz (sign "e")	**-2 to -3**	Weak landing (bad pos./wrong edge/scratching etc)	**-1 to -2**
Unclear ~~wrong~~ edge take off F/Lz (sign "!" or no sign)	**-1 to -2**	Poor take-off	**-1 to -2**
~~Unclear wrong edge take off F/Lz (no sign)~~	~~**-1**~~	Long preparation	**-1 to -2**
SPINS			
SP: Prescribed air position not attained (flying spin)	**-1 to -2**	Poor/awkward, unaesthetic position(s)	**-1 to -3**
Fall	**-3**	Traveling	**-1 to -3**
Touch down with both hands	**-2**	Slow or reduction of speed	**-1 to -3**
Touch down with free foot or one hand	**-1 to -2**	Change of foot poorly done (including curve of entry/exit except when changing direction)	**-1 to -3**
Less than required revolutions	**-1 to -2**		
Incorrect take-off or landing in a flying spin	**-1 to -2**	Poor fly (flying spin/entry)	**-1 to -3**
STEPS			
SP: Listed jumps with more than half rev. included	**-1**	Poor quality of steps, turns, positions	**-1 to -3**
Fall	**-3**	Stumble	**-1 to -2**
Less than half of the pattern doing steps/turns	**-2 to -3**	Does not correspond to the music	**-1 to -2**
CHOREOGRAPHIC SEQUENCES			
Fall	**-3**	Stumble	**-1 to -2**
Inability to clearly demonstrate the sequence	**-2 to -3**	Does not enhance the music	**-1 to -3**
Loss of control while executing the sequence	**-1 to -3**	Poor quality of movements	**-1 to -2**

Figure 2. Original negative rubric items from 2016/2017 grade of execution scoring protocol that deduct points from [2]

23. "traveling during spin"
24. "spin where there is a touch down with both hands"
25. "slow or reduction of speed during spin"
26. "spin where there is a touch down with free foot or one hand"
27. "spin with change of foot poorly done"
28. "spin with less than required revolutions"
29. "incorrect take-off or landing in a flying spin"
30. "flying spin with poor fly"
31. "step sequence with jumps with more than half revolution included"
32. "step sequence with poor quality of steps, turns, positions"
33. "fall during step sequence"
34. "stumble during step sequence"
35. "less than half of the pattern doing steps/turns during step sequence"
36. "step sequence does not correspond to the music"
37. "fall during choreographic sequence"
38. "stumble during choreographic sequence"
39. "inability to clearly demonstrate the sequence during choreographic sequence"
40. "choreographic sequence does not enhance the music"
41. "loss of control while executing the choreographic sequence"
42. "poor quality of movements during choreographic sequence"

1.2. Hand Negatives

Our visual-text pretraining encourages closeness between element embeddings of a particular quality with rubric items of the same quality and repels rubric items of the opposite quality. We find that rubric items that award points ("jump that is unexpected or creative or difficult entry") are not always mutually exclusive with rubric items that lose points ("stepping out of landing in a jump"). Motivated by improved video-language pretraining using synthetic hard negatives [5], we negate each rubric item by hand for a close negative. Since there are only 74 rubric items, this is feasible to do by hand.

List of hand negatives for rubric items (positive):

1. "jump that is expected and uncreative and easy entry"
2. "jump with unclear recognizable steps/free skating movements immediately preceding element"
3. "jump without varied position in the air / delay in rotation"
4. "jump with poor height or distance"
5. "jump with poor extension on landing and creative exit"
6. "jump with poor flow from entry to exit including jump combinations / sequences"
7. "jump that is effortful throughout"
8. "jump not matched to the musical structure"
9. "spin with poor speed or acceleration during spin"
10. "centering a spin slowly"
11. "spin with unbalanced rotations in at least one position"
12. "spin with at most the required number of revolutions"
13. "spin with poor or weak position(s), height, or air/landing position in flying spins"
14. "uncreative or unoriginal spin"
15. "poor control at any phase of spin"
16. "spin not matched to the musical structure"
17. "step sequence with poor energy or execution"
18. "poor speed and acceleration during step sequence"
19. "poorly executed steps during the step sequence"
20. "step sequence without deep clean edges including entry or exit of a turn"
21. "poor control or commitment or inaccurate steps in step sequence"
22. "uncreative or unoriginal step sequence"
23. "effortful throughout step sequence"
24. "step sequence does not enhance the musical structure"
25. "choreographic step sequence with poor flow, energy or execution"
26. "poor speed and acceleration during choreographic step sequence"
27. "choreographic step sequence with poor clarity or precision"
28. "choreographic step sequence with poor control or commitment of whole body"
29. "uncreative or unoriginal choreographic step sequence"
30. "effortful throughout choreographic step sequence"
31. "choreographic step sequence not reflecting concept or character of the program"
32. "choreographic step sequence not enhancing the musical structure"

List of hand negatives for rubric items (negative):

1. "combo contained two jumps"
2. "successfully jump"
3. "completed required preceding steps/movements prior to jump"
4. "correctly rotated jump"
5. "no break between required steps/movements and jump or only 1 step or movement preceding jump"
6. "jump has full rotation including half loop in a combo"
7. "no fall"
8. "jump has good speed, height, distance, air position"
9. "landing on one foot in a jump"
10. "touch down at most one hand in a jump"
11. "not stepping out of landing in a jump"
12. "touch down with no hand and free foot in a jump"
13. "at most one three turns in between in a jump combo"

14. “no loss of flow/direction/rhythm between jumps in combo or sequence”
15. “no or slight wrong edge take off in jump”
16. “jump with strong landing”
17. “jump with clear and right edge take off”
18. “jump with good take-off”
19. “jump with short preparation”
20. “flying spin where prescribed air position is attained”
21. “spin with good/graceful, aesthetic position(s)”
22. “no fall during spin”
23. “in place during spin”
24. “spin where there is no touch down with both hands”
25. “not slow and maintain speed during spin”
26. “spin where there is no touch down with free foot or one hand”
27. “spin with change of foot well done”
28. “spin with at least required revolutions”
29. “correct take-off and landing in a flying spin”
30. “flying spin with good fly”
31. “step sequence with jumps with at most half revolution included”
32. “step sequence with good quality of steps, turns, positions”
33. “no fall during step sequence”
34. “no stumble during step sequence”
35. “at least half of the pattern doing steps and turns during step sequence”
36. “step sequence corresponds to the music”
37. “no fall during choreographic sequence”
38. “no stumble during choreographic sequence”
39. “clearly demonstrate the sequence during choreographic sequence”
40. “choreographic sequence enhances the music”
41. “control while executing the choreographic sequence”
42. “high quality of movements during choreographic sequence”

Margin	MSE (↓)	Sp. Corr. (↑)
0.05	**11.60**	0.801
0.5	12.13	**0.830**
rel	18.43	0.748

Table 1. Visual pretraining margin ablation; reporting performance on Fis-V test after finetuning on train split.

Margin	MSE (↓)	Sp. Corr. (↑)
0.05	16.46	0.782
0.5	16.42	**0.830**
1.0	**10.77**	0.821

Table 2. Visual-text pretraining margin ablation; reporting performance on Fis-V test after finetuning on train split.

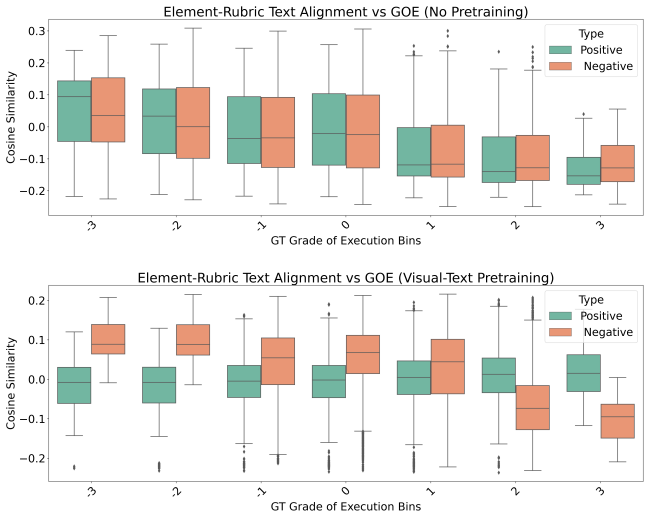

Figure 3. Effect of visual-text pretraining on consistency between element ground truth quality (indicated by GOE on the x-axis) and (1) element embedding-positive rubric text cosine similarity (green) and (2) element embedding-negative rubric text cosine similarity (orange). Cosine similarity is on the y-axis.

2. Experiments

Our experiments and results in this section are:

1. Ablations on the margin hyperparameter for visual-only and vision-text pretraining. We also share quantitative evidence of increased alignment between element embeddings and rubric text of the same quality after vision-text pretraining.
2. Qualitative examples of cross-attention after regularization

2.1. Pretraining Ablation

For both pretraining techniques, we use a triplet loss with a margin hyperparameter, and the distance between pairs is

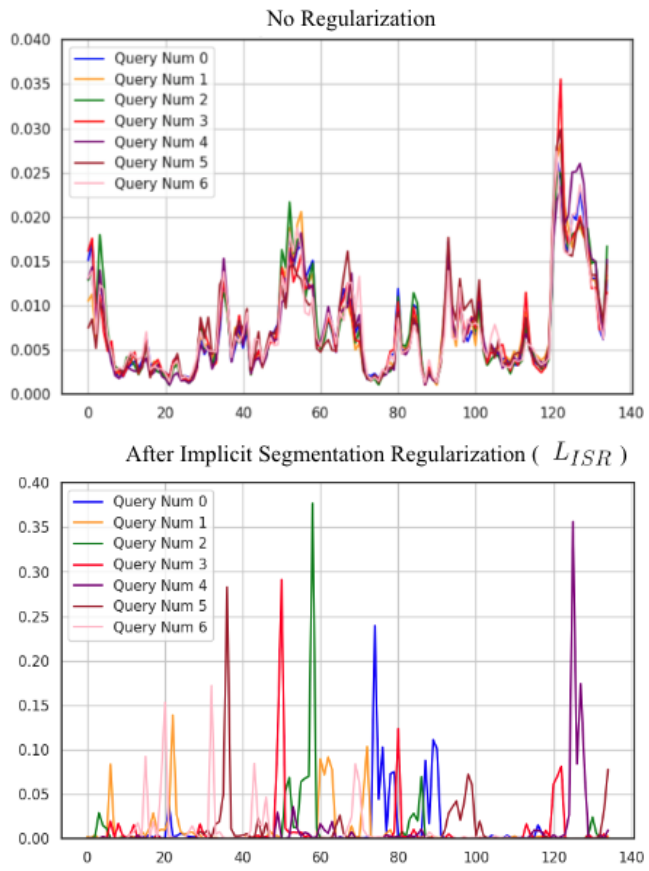

Figure 4. Effect of $\mathcal{L}_{ISR}$ on the cross-attention distribution of each element query and contextualized clip embeddings after the final Transformer decoder layer. Note that attention weights are larger (max 0.4 vs 0.04) in the bottom figure.

calculated using cosine distance. The margin hyperparameter controls the minimum difference in distance between the positive pair and the distance between the negative pair.

In Tab. 1, we observe that the margin has a tradeoff between scoring precision and Spearman correlation, except for relative margin which performs poorly on both metrics.

On visual-text pretraining, we observe that a larger margin (margin=1.0) greatly improves scoring precision in Tab. 2. This is understandable because the metric used in the visual-text triplet loss directly impacts the scoring mechanism. Specifically, the final score is calculated from points associated with each rubric item, weighed by the cosine similarity between element embeddings and rubric items and the cosine distance used in the triplet loss is inversely proportional to cosine similarity. So, a greater degree of alignment of high-scoring element embeddings and positive rubric items is directly beneficial to the scoring task, and for low-scoring element embeddings, vice versa. We see additional evidence of the impact of visual-text pretraining on element-rubric sentiment alignment in Fig. 3; there is higher consistency (e.g., high-scoring elements have higher cosine similarity with rubric text) between the actual element quality and rubric text sentiment after visual-text pretraining compared to no pretraining. Without visual-text pretraining, the average cosine similarity between element embeddings and positive rubric text decreases as the quality (higher GOE) increases, opposite of the desired trend.

2.2. Implicit Segmentation Regularization

Implicit segmentation regularization consists of two regularization losses as mentioned in the main text: orthogonality loss and peak loss. The intuition behind these losses was to apply inductive biases derived from the structure of element segments to the cross-attention maps between element queries and clips. Namely, (1) there is no overlap between element segments, so each element query should attend to different parts of the video, and (2) element segments are contiguous, so queries should naturally focus on a single segment.

When looking at a sample cross-attention map on a test sample in Fig. 4, we observe that all queries in the element transformer have the same cross-attention pattern when no regularization is applied. After implicit segmentation regularization, we observe that the cross-attention map exhibits the structure intended by the regularization. We see that the queries focus on different parts of the video. The cross-attention becomes more concentrated around a smaller number of clips and the peaks are more pronounced, as indicated by higher attention weights.

2.3. Qualitative Results

Next, in Fig. 5, we visualize the clip (peak) that each query attends the most to in the cross-attention map. We observe that queries attend to elements (marked in green), however, some queries will still attend to transitions (marked in red). This may be because step sequences (second row) look similar to transitions, which indicates that context is important to differentiate between a planned step sequence and a transition. While the encoder contextualizes the clip embeddings, it is possible more encoder layers may be needed to model the context better.

We visualize an interesting edge case of our framework in Fig. 6, which demonstrates its ability to predict GOE and handle combination jumps within a short program routine. Typically, combination jumps are graded as a single element by the judges, with the same GOE range as other elements. However, our model seems to be processing a combination jump as two separate elements. In this example, there is a different query attending to each jump in the combination. Specifically, one query focuses on the first jump and the other on the second, each with different qual-

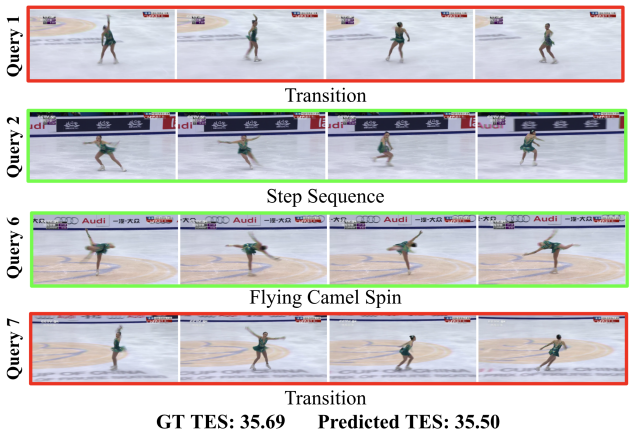

Figure 5. Peak visualization for selected queries and TES prediction for an example video from Fis-V [7]. For each query, we display four frames sampled from the clip with the highest attention (the peak). Clips corresponding to elements are highlighted with a green box, while transitions are marked with a red box.

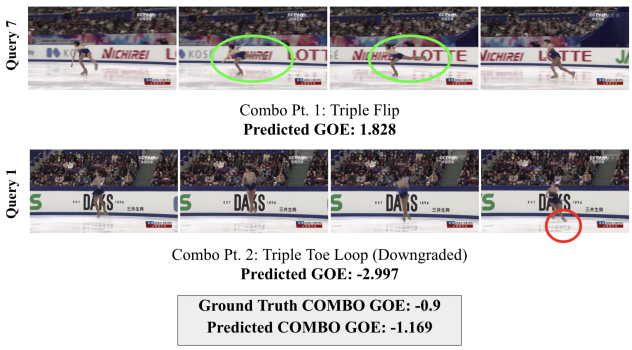

Figure 6. Visualization of GOE prediction edge case. In a short program routine, there are typically 7 elements, one of which must be a combination jump (i.e., a combination of two jumps). Although the judges grade this as a single element, with the same GOE range as other elements, our model treats it as two separate elements. In the example above, the model compensates for this by using two different queries to focus on each jump in the combination. Specifically, Query 7 focuses on the first jump, which has a solid landing, and assigns it a high score. Meanwhile, Query 1 focuses on the second jump, which is under-rotated and poorly landed, resulting in a lower score. If we manually combine these scores (predicted COMBO GOE), it's fairly close to the ground truth. The green and red circles are for illustrative purposes to guide the reader's attention to quality-indicative regions.

ity assessments. The predicted GOE for these individual jumps, when considered together (-1.2), closely align with the overall ground truth GOE (-0.9) for the combination element. While it's not ideal for two queries to point to the same element, this shows the potential of the model to handle complexity within our framework, and it is promising that both parts were accurately scored.

Method	MSE (↓)	Sp. Corr. (↑)
MLP-Mixer* [6] (2023)	81.24	0.880
SGN* [3] (2024)	79.08	0.890
Base Value Lookup	72.94	0.943
Element Transformer + RCS	43.88	0.947
+ $\mathcal{L}_{ISR}$	**38.91**	**0.948**

Table 3. Evaluation on FS800 [6] validation subset. Note our results are without pretraining. RCS = Rubric-Constrained Scoring. * indicates that results are not directly comparable due to differences in experimental protocol.

2.4. FS1000 Evaluation

We evaluate on the newer FS800 dataset (a subset of FS1000), which includes a more diverse set of programs, featuring both men's and women's skating, unlike Fis-V [7], which focuses solely on women's solo short programs. Acquiring the FS1000 dataset was not trivial, as it is not publicly available (Baidu). Due to significant effort in data cleaning and navigating non-intuitive folder naming conventions, we ended up evaluating on a subset of the data, referred to as FS800. We excluded ice dancing and pairs events from this subset due to their different scoring protocols. We trained on 400 videos and evaluated on 86 videos. While this dataset is roughly the same size as Fis-V, there is no overlap between the two. We use pre-extracted Timesformer [1] clip embeddings since it is already provided by the dataset [6], instead of Video Swin [4].

In Tab. 3, we evaluate the performance of our models on the FS800 subset. First, the high baseline performance of Base Value Lookup (BVL) is aligned with Table 1 results in the main paper. When comparing these two tables, it seems unusual that mean-squared error is high on FS800, however, this is because of long programs in the dataset which have a larger score range. Prior results on FS1000 (first two rows), also show high MSE. Similar to the results on Fis-V, we show improved performance with our full method (without pretraining) compared to BVL. We also show the benefit of applying our implicit segmentation regularization on top of the Element Transformer with the Rubric-Constrained Scoring head.

While our method significantly reduces mean-squared error, it only slightly improves score ranking compared to BVL. Limited improvement on score ranking may be due to the current trend of athletes being rewarded more for difficulty, with those attempting harder programs receiving the highest scores. Nevertheless, we show improvements in both metrics compared to using only Base Value Lookup, validating that our framework has increased *quality* estimation capability on this dataset despite the increased complexity of this dataset.

References

[1] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In *Proceedings of the International Conference on Machine Learning (ICML)*, July 2021. 7

[2] Ottavio Cinquanta and Fredi Schmid. ISU Communication No. 2000 (revised June 14, 2016): SINGLE & PAIR SKATING - Scale of Values, Levels of Difficulty and Guidelines for marking Grade of Execution, season 2016/17, April 14 2016. 2, 3

[3] Zexing Du, Di He, Xue Wang, and Qing Wang. Learning semantics-guided representations for scoring figure skating. *IEEE Transactions on Multimedia*, 26:4987–4997, 2024. 7

[4] Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. Video swin transformer. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3192–3201, 2021. 7

[5] Liliane Momeni, Mathilde Caron, Arsha Nagrani, Andrew Zisserman, and Cordelia Schmid. Verbs in action: Improving verb understanding in video-language models. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15533–15545, 2023. 4

[6] Jingfei Xia, Mingchen Zhuge, Tiantian Geng, Shun Fan, Yuantai Wei, Zhenyu He, and Feng Zheng. Skating-mixer: Long-term sport audio-visual modeling with mlps. In *AAAI Conference on Artificial Intelligence*, 2022. 7

[7] C. Xu, Yanwei Fu, Bing Zhang, Z. Chen, Yu-Gang Jiang, and X. Xue. Learning to score figure skating sport videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 30:4578–4590, 2018. 7