SACA Net: Cybersickness Assessment of Individual Viewers for VR Content via Graph-based Symptom Relation Embedding - *Supplementary Material* -

1 Network Details

Table 1. Network details of the visual expectation generator in the sensory mismatch detector of the stimulus symptom context guider.

Visual Expectation Generator		
Layer	**Filter/ Stride**	**Output Size ($h\times w\times c$)**
ConvLSTM1	3×3/ (2, 2)	112×112×16
ConvLSTM2	3×3/ (2, 2)	56×56×32
ConvLSTM3	3×3/ (2, 2)	28×28×64
ConvLSTM4	3×3/ (2, 2)	14×14×128
DeConvLSTM1	3×3/ (2, 2)	28×28×64
DeConvLSTM2	3×3/ (2, 2)	56×56×32
DeConvLSTM3	3×3/ (2, 2)	112×112×16
DeConvLSTM4	3×3/ (2, 2)	224×224×3

Table 2. Network details of the mismatch encoder and the visual encoder in the stimulus symptom context guider. The mismatch encoder and the visual encoder have the same network structure but do not share their weights.

Mismatch Encoder / Visual Encoder		
Layer	**Filter/ Stride**	**Output Size ($l\times h\times w\times c$)**
3D-Conv1	3×3×3/ (1, 2, 2)	9×112×112×8
3D-Conv2	3×3×3/ (1, 2, 2)	7×56×56×16
3D-Conv3	3×3×3/ (1, 2, 2)	5×28×28×32
3D-Conv4	3×3×3/ (1, 2, 2)	3×14×14×64
3D-Conv5	3×3×3/ (1, 2, 2)	1×7×7×64

Table 3. Network details of the global context encoder and the symptom group feature extraction in the stimulus symptom context guider. From FC2, it is applied individually for each symptom group. The feature after FC3 indicates the symptom group feature.

Global Context Encoder / Symptom Group Feature Extraction		
Layer	**Filter/ Stride**	**Output Size ($h \times w \times c$)**
2D-Conv	3×3/ (1, 1)	7×7×256
FC1	64/ (-)	1×1×64
FC2	32/ (-)	1×1×32
FC3	32/ (-)	1×1×32
FC4	1/ (-)	1×1×1

Table 4. Network details of the time-wise encoder, the frequency band attention, the time-freq-wise encoder, and the symptom feature extraction in the physiological symptom guider. From FC4_2, it is applied individually for each symptom. The feature after FC4_3 indicates the symptom feature.

Time-wise Encoder/Frequency Band Attention /Time-Freq-wise Encoder/Symptom Feature Extraction		
Layer	**Filter/ Stride**	**Output Size ($h \times w \times c$)**
1D-Conv1_1	1×3/ (1, 1)	48×128×32
1D-Conv1_2	1×3/ (1, 1)	48×128×32
1D-Conv1_3	1×3/ (1, 2)	48×64×32
2D-Conv2_1	3×3/ (2, 2)	24×32×16
2D-Conv2_2	3×3/ (2, 2)	12×16×8
2D-Conv2_3	3×3/ (2, 2)	6×8×1
FC2_1	5/ (-)	1×1×5
2D-Conv3_1	3×3/ (1, 1)	48×64×32
2D-Conv3_2	3×3/ (2, 1)	24×64×32
2D-Conv3_3	3×3/ (2, 2)	12×32×32
2D-Conv4_1	3×3/ (1, 1)	12×8×32
2D-Conv4_2	3×3/ (2, 2)	6×4×32
FC4_1	256/ (-)	1×1×256
FC4_2	16/ (-)	1×1×16
FC4_3	16/ (-)	1×1×16
FC4_4	4/ (-)	1×1×4
FC4_5	1/ (-)	1×1×1

2 Network Computational Cost

Table 5. Weight parameter size and inference time for the proposed network. The inference time is checked based on a single TITAN XP GPU.

Model	# of Weights	Inference Time (sec)
Stimulus Symptom Context Guider	3.03M	1.646
Physiological Symptom Guider	0.16M	0.057
Symptom Relation Embedder	0.04M	0.016
Total	**3.23M**	**1.719**

3 Additional Quantitative Results

Table 6. Total SSQ score prediction performances according to the mismatch feature from the sensory mismatch detector in the stimulus symptom context guider.

VRSA DB-Shaking				VRSA DB-FR			
Method	PLCC	SROCC	RMSE	Method	PLCC	SROCC	RMSE
Proposed Method (w/o Mismatch Features)	0.725	0.606	30.069	Proposed Method (w/o Mismatch Feature)	0.777	0.640	24.992
Proposed Method (w/ Mismatch Feature)	**0.751**	**0.679**	**25.373**	**Proposed Method (w/ Mismatch Feature)**	**0.801**	**0.671**	**22.937**

Table 7. Total SSQ score prediction performances according to Relational Embedding (RE) and Stimulus Context (ST).

VRSA DB-Shaking				VRSA DB-FR			
Method	PLCC	SROCC	RMSE	Method	PLCC	SROCC	RMSE
Proposed Method (w/o RE, w/o ST)	0.610	0.461	35.070	Proposed Method (w/o RE, w/o ST)	0.745	0.599	25.611
Proposed Method (w/ RE, w/o ST)	0.677	0.544	32.961	Proposed Method (w/ RE, w/o ST)	0.768	0.619	24.631
Proposed Method (w/ RE, w/ ST)	**0.751**	**0.679**	**25.373**	**Proposed Method (w/ RE, w/ ST)**	**0.801**	**0.671**	**22.937**

4 Learning Curves

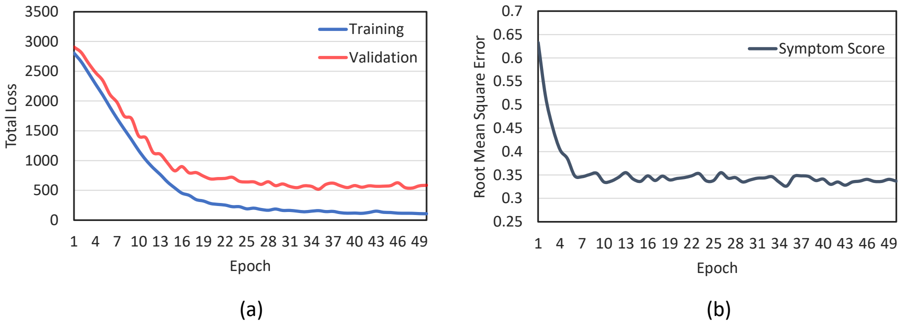

Fig. 1. (a) Total loss for training and validation according to the learning epoch. (b) Root mean square error (RMSE) validation for symptom score according to the learning epoch. Both results are from one training fold on VRSA DB-FR.