Supplementary Materials: All rivers run into the sea: Unified Modality Brain-like Emotional Central Mechanism

Anonymous Authors

1 OVERVIEW

The additional material mainly consists of two parts: the research motivation and the experiment.

- 2 Research Motivation is a supplement to 1 Introduction part,
- 3 Experimental part is a supplement to 4 Experiment.
- 4 Future Work part is a supplement to 5 Conclusion

2 RESEARCH MOTIVATION

Cross-modal plasticity refers to the brain's ability to adjust and reorganize its functions to enhance the processing capabilities of other sensory modalities following the loss or absence of input from a particular sensory modality[3, 10, 11]. This phenomenon primarily involves neural plasticity, which is the structural and functional changes in neurons and neural networks under the influence of experience or the environment[1, 12]. These changes include the formation, strengthening, or weakening of synapses, and the establishment of new neural pathways[4, 8]. In cases of sensory deprivation, such as blindness or deafness, the brain regions originally receiving input from the now-absent modality no longer do so. Instead, neural centers of other modalities may take over these areas, reallocating and optimizing their functions. For example, in experiments involving visual deprivation and auditory enhancement, the brain areas originally processing visual information, such as the visual cortex, begin to process auditory or tactile information[7]. This reorganization can lead to increased auditory and tactile sensitivity, as observed in the extraordinary auditory localization abilities of blind individuals. Similarly, in experiments involving auditory deprivation and visual/tactile enhancement, individuals with hearing loss exhibit enhanced visual and tactile functions, such as improved visual-spatial processing abilities and heightened tactile perception[9]. The mechanisms underlying cross-modal plasticity involve a variety of molecular and cellular events, such as changes in neurotrophic factors, synaptic connectivity reorganization, and the rebalancing of inhibitory and excitatory neurotransmissions. The brain utilizes these mechanisms to optimize the remaining sensory inputs to compensate for the lost senses. This plasticity helps individuals adapt to sensory loss and enhances the functions of other senses, revealing the brain's remarkable adaptability to changes in sensory inputs.

Neuroanatomy is the scientific field that studies the structure and function of the brain, including how the brain processes information through its complex network structures[2, 14]. Specifically, in the context of multimodal information processing, neuroanatomy demonstrates how the brain integrates information from our various sensory systems, including vision, hearing, touch, smell, and taste[6, 15]. The parietal lobe, located in the upper part of the brain, is a primary area for processing tactile, visual, and spatial information[2, 5]. The temporal lobe primarily handles auditory information and some visual memory. It contains several key multimodal areas, such as the superior temporal sulcus (STS), which is an important area for integrating visual and auditory information related to facial expressions and body movements, as well as associated sounds[13]. Additionally, the temporal lobe is involved in language comprehension and emotional processing. The frontal lobe, located at the front of the brain, is central to decision-making, planning, and social behavior. The dorsolateral prefrontal cortex (DLPFC) within the frontal lobe is key for multimodal information processing, responsible for merging information from different sensory sources to support complex cognitive tasks such as problem-solving and decision-making. The insula, deep within the brain, acts as an integration center for information from multiple sensory systems, playing a critical role in regulating emotional responses, pain perception, taste, smell, and visceral sensations. Its role in multimodal information processing includes integrating internal and external sensory signals and assessing emotionally relevant stimuli. The amygdala, a critical area in the brain for processing emotional responses, especially in managing emotions like fear and happiness, receives and integrates various sensory information, such as visual and auditory signals that relate to emotional responses and social prompts. Multimodal information processing relies not only on the functions of individual brain regions but also on extensive neural networks. These networks are interconnected through complex axonal connections. For example, visual-auditory information is integrated in the superior temporal sulcus and other areas of the temporal lobe, which have extensive connections with the parietal and frontal lobes, jointly participating in tasks such as spatial localization and speech understanding.

3 EXPERIMENTAL

In the experimental section, regarding the performance experiments, the confusion matrix for Table 1 has already been presented in the main text. We have additionally included the confusion matrices for Tables 2, 3, and 4 as Figures 1, 2, 3 and 4 in the supplementary materials. In these experiments, SENET consistently demonstrated performance as shown in the tables, and even exceeded the performance indicated therein.

4 FUTURE WORK

In the field of affective computing, leveraging information from various sensory channels is crucial for a nuanced understanding and interpretation of human emotions. Inspired by the theory of cross-channel plasticity, we propose a new unified channel paradigm for affective computing, named UMBEnet. UMBEnet seamlessly integrates multimodal information across visual, auditory, and textual domains to create a more proficient system that enhances the accuracy and resilience of emotion recognition efforts. Our approach is grounded in the complex neuroanatomical structures of the human brain, incorporating a brain-like emotional processing framework that utilizes inherent cues and dynamic cue pools, along with sparse feature fusion techniques. Rigorous experimental validation on the most scalable benchmarks in the DFER field—specifically, DFEW, FERV39k, and MAFW—has definitively confirmed that UMBEnet surpasses the performance benchmarks set by existing state-of-the-art (SOTA) methods. Particularly in scenarios characterized by multi-channel configurations or the absence of such configurations, UMBEnet clearly exceeds contemporary leading solutions.

We believe that the BEPF framework, DS structure, and SFF module proposed in this paper offer significant guidance for addressing issues of fusion and modality absence within the multimodal domain. Our work provides the multimodal community with a novel and distinct model framework for multimodal fusion and addressing modality absence, differing from previous approaches. In the future, we aim to extend the BEPF framework to more multimodal domains, believing it to be a transferable multimodal framework that requires minimal modification for adaptation.

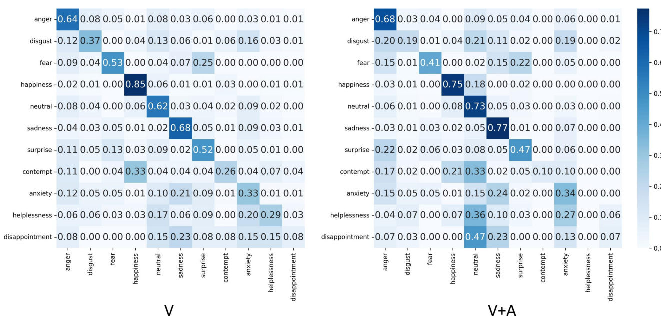

Figure 1: Confusion Matrices of Overall Model Performance Comparison (UMBEnet vs. other SOTA methods on MAFW for 11-class classification)

REFERENCES

[1] Daphne Bavelier and Helen J Neville. 2002. Cross-modal plasticity: where and how? *Nature Reviews Neuroscience* 3, 6 (2002), 443–452.

[2] Yong Fan, Feng Shi, Jeffrey Keith Smith, Weili Lin, John H Gilmore, and Dinggang Shen. 2011. Brain anatomical networks in early human brain development. *Neuroimage* 54, 3 (2011), 1862–1871.

[3] Hannah Glick and Anu Sharma. 2017. Cross-modal plasticity in developmental and age-related hearing loss: Clinical implications. *Hearing research* 343 (2017), 191–201.

[4] Drew B Headley and Norman M Weinberger. 2015. Relational associative learning induces cross-modal plasticity in early visual cortex. *Cerebral Cortex* 25, 5 (2015), 1306–1318.

[5] Gregory Hickok. 2009. The functional neuroanatomy of language. *Physics of life reviews* 6, 3 (2009), 121–143.

[6] Stanley Jacobson, Elliott M Marcus, and Stanley Pugsley. 2011. *Neuroanatomy for the Neuroscientist.* Springer.

[7] Teijo Kujala, K Alho, M Huotilainen, RJ Ilmoniemi, Anne Lehtokoski, A Leinonen, T Rinne, OC Salonen, J Sinkkonen, C-G Standertskjöuld-Nordenstam, et al. 1997. Electrophysiological evidence for cross-modal plasticity in humans with early- and late-onset blindness. *Psychophysiology* 34, 2 (1997), 213–216.

[8] Ron Kupers and Maurice Ptito. 2014. Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neuroscience & Biobehavioral Reviews* 41 (2014), 36–52.

[9] Lindsay B Lewis, Melissa Saenz, and Ione Fine. 2010. Mechanisms of cross-modal plasticity in early-blind subjects. *Journal of neurophysiology* 104, 6 (2010), 2995–3008.

[10] Stephen G Lomber, M Alex Meredith, and Andrej Kral. 2010. Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nature neuroscience* 13, 11 (2010), 1421–1427.

[11] Maurice Ptito and Ron Kupers. 2005. Cross-modal plasticity in early blindness. *Journal of integrative neuroscience* 4, 04 (2005), 479–488.

[12] K Sathian and Randall Stilla. 2010. Cross-modal plasticity of tactile perception in blindness. *Restorative neurology and neuroscience* 28, 2 (2010), 271–281.

[13] Richard S Snell. 2010. *Clinical neuroanatomy.* Lippincott Williams & Wilkins.

[14] Charles Watson, Matthew Kirkcaldie, and George Paxinos. 2010. *The brain: an introduction to functional neuroanatomy.* Academic Press.

[15] Stuart Zola-Morgan and Larry R Squire. 1993. Neuroanatomy of memory. *Annual review of neuroscience* 16, 1 (1993), 547–563.

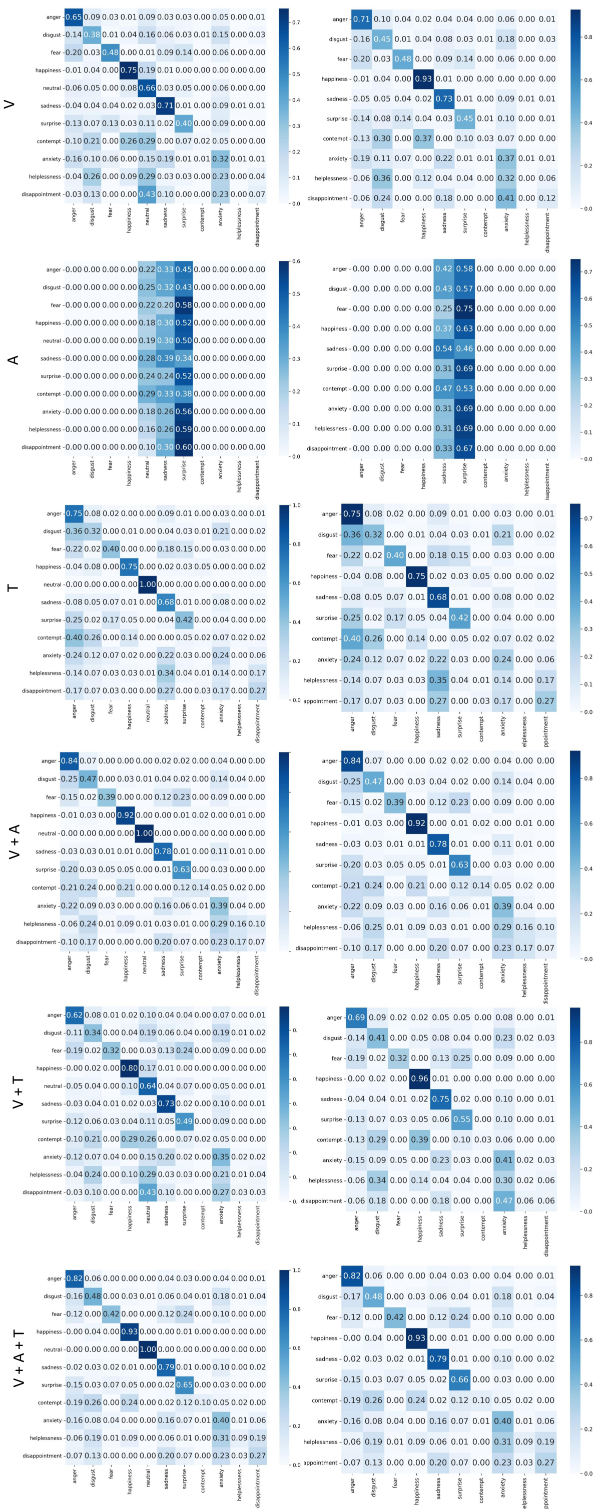

Figure 3: Confusion Matrices of Missing Mode (UMBEnet on MAFW for 11-class and 10-class classification).

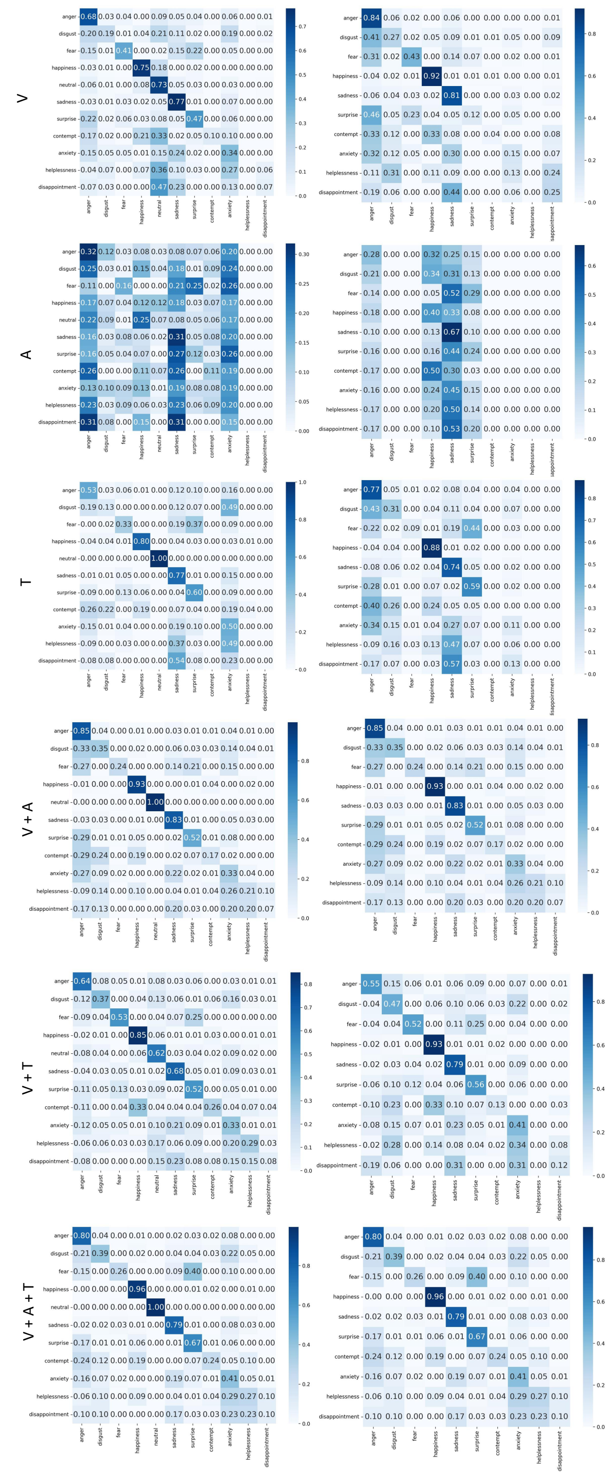

Figure 2: Confusion Matrices of Multimode Performance Comparison (UMBEnet on MAFW for 11-class and 10-class classification).

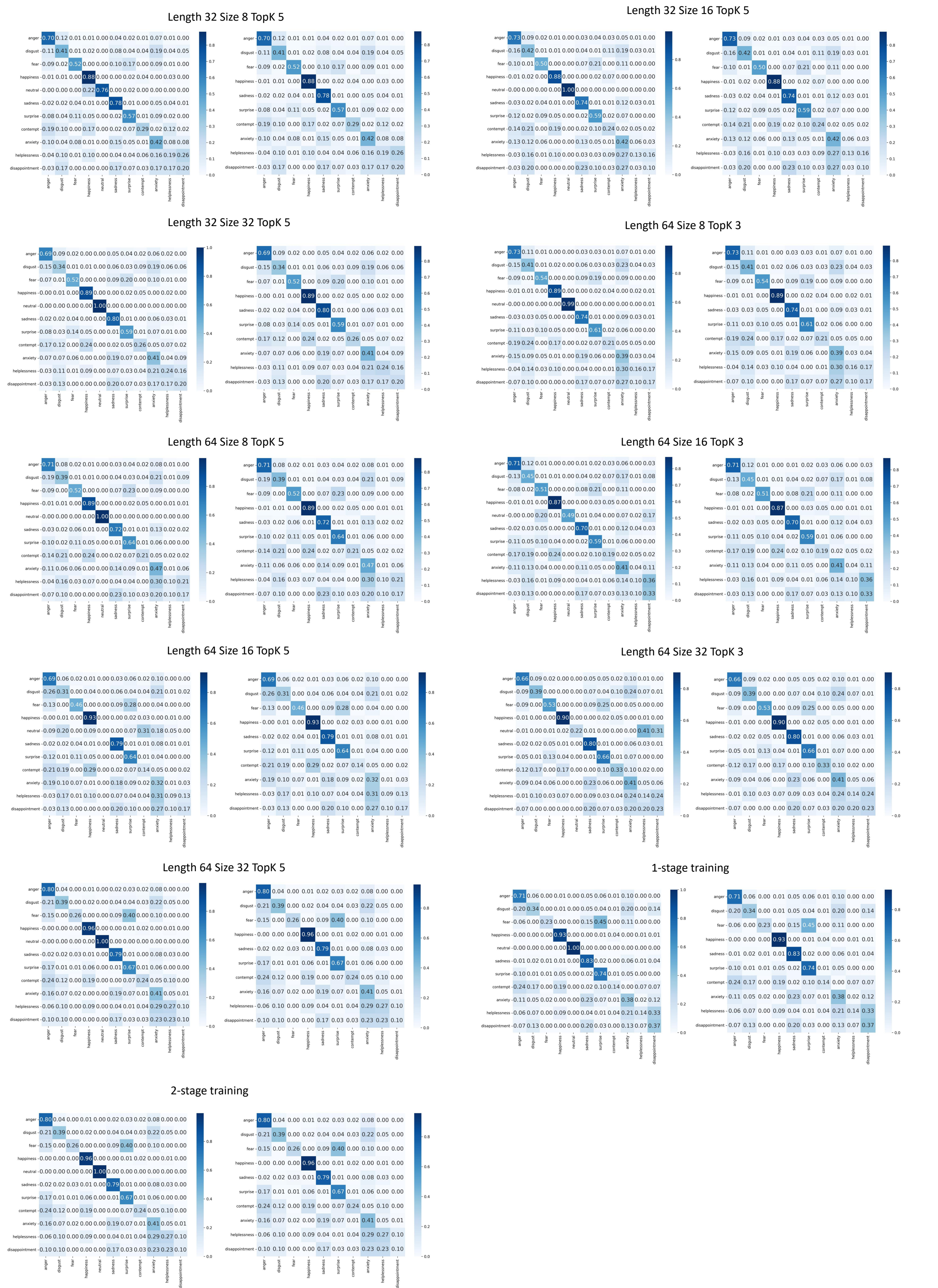

Figure 4: Confusion Matrices of UMBEnet Hyperparameter Ablation Study (UMBEnet on MAFW for 11-class and 10-class classification).