Abstract

Facial emotion recognition (FER) is now a key component of human-computer interaction systems, but existing techniques are hindered by inaccuracies, computational complexities, and performance variability in diverse environmental conditions. This thesis presents a detailed study of high-level ensemble learning architectures to alleviate these intrinsic constraints via three novel deep learning paradigms.

The paper suggests three synergistic approaches collectively promoting the building of facial emotion recognition (FER) systems. M^{3}SI-Net initially describes a new Inception block-based fusion structure and a new Re-parameterized Swish1 activation function, promoting improved feature representation through multi-scale convolutional operations. The Re-parameterized Swish1 activation function demonstrates better gradient flow properties compared to conventional activation functions, with better training stability and convergence rates. Empirical verification on benchmarking datasets (CK+, JAFFE) demonstrates accuracy improvements compared to baselines with computational tractability.

The second innovation, SIG-Net, proposes an adopted sigmoid-based ensemble network with an integration of different learning paradigms to promote robustness. The novel sigmoid function includes adaptive scaling parameters to improve activation properties for emotion-specific features. The ensemble model shows improved generalization ability in cross-dataset testing with average accuracy improvements compared to single base learners.

The third contribution solves the urgent need for lightweight FER systems via an entropy-based deep feature selection algorithm and efficient KNN classification. Utilizing information gain principles, the method accomplishes dimensionality reduction with the classification accuracy still at a low, relative to full-feature models. The method supports real-time emotion recognition on resource-limited devices with inference time decreased relative to state-of-the-art deep learning methods.

Detailed comparative evaluation indicates complementary advantages in the three methods: M3SI-Net is superior in accuracy-driven applications, the light framework allows mobile deployment, and SIG-Net offers best robustness for harsh environments. The work sets new performance standards on standard FER benchmarks and offers practical deployment advice for various application environments.

The breakthroughs go beyond architectural innovations to encompass theoretical innovations in ensemble learning theory, activation function design, and information theory-based feature selection. These results have significant impli-

cations for future emotion recognition systems utilized in autonomous vehicles, health monitoring, learning systems, and social robotics systems, thus paving the way for more reliable and efficient human-machine emotional interfaces.