Deep Learning | Final Project Proposal

Topic: Natural Language Processing (sentiment analysis - emotion recognition)

Dataset: FER2013 (Facial Expression Recognition 2013 Dataset)

Emotion recognition from facial expressions addresses the urgent need for affective computing in socially assistive robotics, a field increasingly vital for supporting vulnerable groups such as the elderly and children with neurodevelopmental conditions. The FER2013 dataset, introduced by Goodfellow et al. (2013) as part of an ICML challenge, serves as a foundational benchmark for this task, offering 35,887 grayscale images labeled with seven emotions (anger, disgust, fear, happiness, sadness, surprise, neutral). While FER2013 is relatively small for training deep networks from scratch, posing challenges like class imbalance (e.g., limited "disgust" samples) and overfitting, it remains widely adopted due to its real-world variability (lighting, pose, occlusion) and standardized evaluation. Researchers such as Minaee et al. (2021) and Zeng et al. (2022) have demonstrated that techniques like transfer learning (using models pretrained on larger datasets like ImageNet) and data augmentation (e.g., rotations, flips) can mitigate dataset limitations, enabling robust model performance.

To address FER2013's challenges, this project will use a pretrained ResNet50 architecture, customized for emotion recognition. While ResNet50 is a standard network for image tasks, its original form (designed for 224x224 RGB images and 1,000-class ImageNet labels) requires adaptation. First, grayscale FER2013 images (48x48 pixels) will be upsampled and replicated across three channels to match ResNet50's input dimensions. The final classification layer will be replaced with a 7-neuron dense layer (for the seven emotions) and augmented with dropout layers to reduce overfitting. Additionally, initial layers of the pretrained model will be frozen to preserve generic feature extraction (edges, textures), while later layers will be fine-tuned on FER2013 to capture emotion-specific patterns. This hybrid approach, inspired by Cai et al. (2021), balances computational efficiency and task specificity, ensuring the model generalizes to real-world scenarios like interpreting subtle facial cues in elderly care or ASD therapy.

This capability is critical for applications in social robotics, where systems must interpret nuanced emotions to assist elderly individuals facing loneliness or children with autism spectrum disorder (ASD). For instance, robots like PARO (Wada et al., 2005) and NAO (Tapus et al., 2007) rely on accurate emotion recognition to deliver empathetic interactions, fostering trust and engagement. By refining models on FER2013, researchers advance technologies that can dynamically adapt to user affect, aligning with healthcare goals such as reducing caregiver burden and enhancing personalized support, as highlighted in reviews on socially assistive robotics (Broekens et al., 2009). Thus, while FER2013's size necessitates careful methodology, its role in bridging technical innovation and societal impact remains indispensable.

To operate these advancements, TensorFlow/Keras was selected as the framework due to its high-level APIs, which streamlines the implementation of transfer learning, which is a necessity for adapting pretrained architectures like ResNet50 to the FER2013 dataset. Customizations include resizing grayscale images (48x48 pixels) to match ResNet50's input dimensions and integrating dropout layers to mitigate overfitting, a common challenge with limited datasets. The framework's ImageDataGenerator further facilitates data augmentation (e.g., rotations, flips), enhancing robustness to real-world variability such as lighting and pose. Keras's intuitive interface significantly reduces development time, allowing researchers to prioritize constructing models tailored to identifying the seven target emotions (anger, disgust, fear, happiness, sadness, surprise, neutral). Additionally, TensorFlow/Keras offers extensive community support and documentation, accelerating iterative development and enabling rapid prototyping. While alternatives like PyTorch were considered, TensorFlow/Keras prioritizes deployment readiness, with tools like TensorFlow Lite enabling real-time inference on edge devices, an essential feature for scalable social robotics applications. By balancing computational efficiency with expressive power, this framework ensures the model's practical utility in contexts such as elderly companionship or ASD therapy, where real-time, adaptive emotional intelligence is paramount.

To inform the technical implementation of ResNet50 for emotion recognition on the FER2013 dataset, this project draws upon foundational research such as Goodfellow et al.'s analysis of representation learning challenges and advanced methodologies like partial facial landmark integration. Open-source resources, including TensorFlow tutorials and GitHub repositories (e.g., FER-2013-CNN), provide practical insights into architectural adaptations for the seven target emotions. Further guidance will be derived from François Chollet's Deep Learning with Python, which elucidates Keras-based optimization strategies to enhance real-world applicability, which is critical for deploying models in socially assistive contexts, such as robots aiding elderly individuals or children with autism spectrum disorder (ASD).

The network's performance will be evaluated using metrics tailored to FER2013's unique constraints. Accuracy and F1-score (prioritized to address class imbalances, particularly for underrepresented emotions like "disgust") will quantify overall efficacy, while a confusion matrix will identify systematic misclassifications (e.g., confusion between "fear" and "surprise"). To mitigate overfitting risks inherent to small datasets, 5-fold cross-validation will monitor validation loss across diverse splits, ensuring robustness. These evaluations collectively ensure the model's reliability in social robotics applications, where precise emotion interpretation is paramount for fostering trust in vulnerable populations, such as elderly users experiencing isolation or children with ASD requiring nuanced emotional support.

Estimated Timeframe:

The project will follow a structured timeline to ensure methodological rigor and alignment with social robotics objectives. Below is the proposed schedule for development and evaluation:

Phase	**Tasks**	**Duration**
Data preparation	Preprocess images (resize, normalize, augment), split into train/val/test	1 week
Model setup	Customize pretrained network (input layer, dropout), compile with Adam	4 days
Training & Tuning	Train with augmentation, hyperparameter tuning (learning rate, epochs)	2 weeks
Evaluation	Calculate metrics (F1, confusion matrix), compare with baselines	3 days
Final report	Document results, visualize attention maps, summarize findings	3 days

References:

1. I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner et al., "Challenges in representation learning: A report on three machine learning contests," arXiv:1307.0414, 2013.
2. J. Cai, Y. Meng, Z. Chen, S. Lu, and A. K. Nandi, "LTNet: A learned tokenization transformer for facial expression recognition," IEEE Trans. Affect. Comput., early access, 2021, Art. no. 9669082, doi: 10.1109/TAFFC.2021.9669082.
3. J. Zeng, S. Shan, and X. Chen, "Facial expression recognition with swin transformer," IEEE Trans. Image Process., vol. 31, pp. 2001–2014, 2022, doi: 10.1109/TIP.2022.3148872.
4. S. Minaee, M. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Going deeper in facial expression recognition using deep neural networks," IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2021, pp. 1–10, doi: 10.1109/WACV48630.2021.00009.
5. A. Tapus, M. J. Matarić, and B. Scassellati, "The role of physical embodiment in human-robot interaction," IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN), 2007, pp. 553–558, doi: 10.1109/ROMAN.2007.4415209.
6. K. Wada, T. Shibata, T. Saito, and K. Tanie, "Robot therapy for elderly residents in a care house: Effects of physical and social interaction," IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN), 2005, pp. 581–586, doi: 10.1109/ROMAN.2005.1513855.
7. J. Broekens, M. Heerink, and H. Rosendal, "Assistive social robots in elderly care: A review," Gerontechnology, vol. 8, no. 2, pp. 94–103, 2009, doi: 10.4017/gt.2009.08.02.002.00.
8. F. Chollet, Deep Learning with Python. Shelter Island, NY, USA: Manning Publications, 2018.