Analysis of Open-Unmix Model Experimental Study

Research Questions

The study focused on two main questions: 1. Understanding how batch size, hidden size, and samples per track affect the model's loss and training time 2. Exploring if modifications to the bidirectional LSTM could enable real-time audio separation without significant quality loss

Experimental Design

Hypotheses

1. Batch size and hidden size increases would reduce model loss, with non-linear training time relationships
2. Replacing bidirectional LSTM with unidirectional LSTM could achieve real-time separation
3. Hidden size has a direct relationship with loss convergence up to a certain point, after which it becomes inversely proportional

Variables

1. **Independent Variables**
    - Number of epochs
    - Number of parallel workers
    - Batch size
    - Hidden size
    - Samples per track
    - LSTM directionality
2. **Control Variables**
    - Patience parameters
    - Learning rate decay settings
    - Frequency bandwidth
    - Number of audio channels
3. **Dependent Variables**
    - Training and validation loss
    - Training time
    - Model accuracy on validation set

Experimental Results

Model Variations and Performance

1. **Base Model (open-unmix)**
    - Batch size: 8
    - Hidden size: 256

- Samples per track: 16
- Loss: 9.536
- Training time: 7:50 (1 epoch)

2. **open-unmix2**
   - Reduced batch size to 4
   - Loss improved to 8.55
   - Similar training time: 7:43
3. **open-unmix3**
   - Increased batch size to 16
   - Doubled hidden size to 512
   - Loss increased to 11.2
   - Training time: 7:42
4. **open-unmix4**
   - Same settings as open-unmix3
   - Extended training to 4 epochs
   - Loss reduced to 3.2
   - Training time increased to 31:09:00
5. **open-unmix5**
   - Batch size: 32
   - Hidden size: 512
   - Samples per track: 32
   - Achieved lowest loss: 2.26
   - Training time: 1:18:00 (5 epochs)

Key Findings

1. **Batch Size Impact**
   - No straightforward relationship between batch size and loss reduction
   - Training time not directly related to batch size due to proportional batch number reduction
2. **Hidden Size Effects**
   - Optimal hidden size found to be 256 for 16kHz signals
   - Larger hidden sizes led to potential overfitting
   - ReLU activation layers helped maintain model robustness with increased hidden size
3. **Real-time Processing**
   - Unidirectional LSTM showed potential for real-time processing
   - Performance heavily dependent on hardware capabilities
   - 16kHz frequency cropping significantly improved processing speed

Conclusions

1. **Training Parameters**
   - Complex relationship between batch size and model performance
   - Training time more dependent on epochs than batch size
   - Hidden size optimization crucial for preventing overfitting

2. **Real-time Processing**
   - Achievable with unidirectional LSTM
   - Success dependent on hardware capabilities
   - Frequency cropping helps achieve faster processing
3. **Model Architecture**
   - 256 hidden size optimal for 16kHz signals
   - ReLU activation important for model stability
   - Trade-off between processing speed and output quality

Future Directions

- Further optimization of unidirectional LSTM for real-time applications
- Exploration of different frequency cropping resolutions
- Potential applications in commercial music processing systems like Apple Music