Medical Imaging

Team VcLD

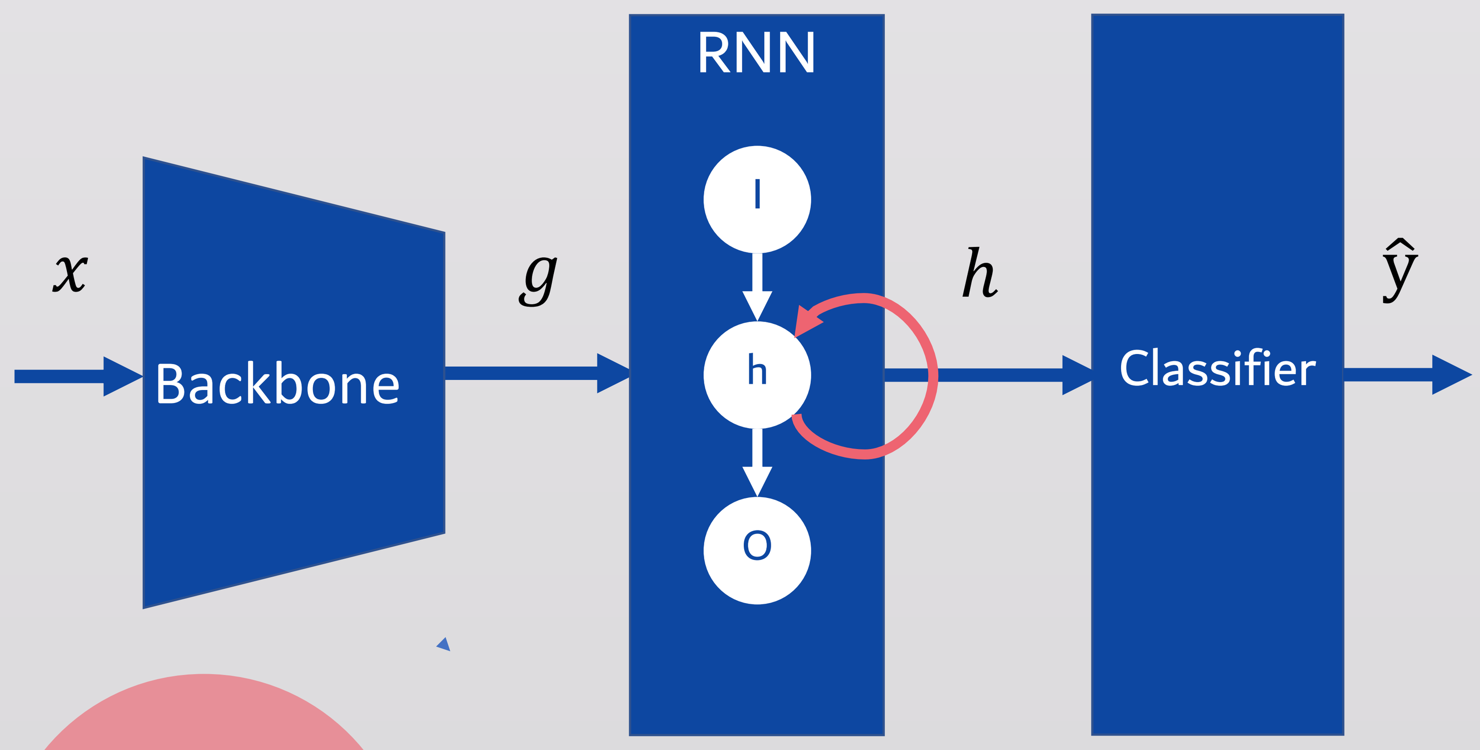
R08922030 Zi-Heng Chen
R08521605 Ting-Yu Dai
R08942067 Shao-Yeh Huang
R08725022 Po-Lin Lai

Data Preprocessing

- Random rotation
- Horizontal / Vertical flip
- Brightness adjustment
- MultiScaleCrop ... etc

Model Architecture

Backbone + RNN + Classifier
ResNet18 + BiGRU + FC



Hyperparameters

- Batch Size: 4 (patient_ids)
- Optimizer: AdamW
- γ_-, γ_+ = 4, 1
- T = 8
- $\epsilon = 0.5$
- p_1, p_2 = 15, 0.15
- Learning Rate: 0.001
- Learning Rate Schedule: Reduce the learning rate by half each time the validation loss is non-decreasing for three continuous epochs. Stop the training stage if learning rate is smaller than 1e-6.

Performance

F2-score evaluation:

Leaderboard	Public	Private
Single Model	0.78255	0.78312
Ensemble Model	0.79395	0.79353
Small dataset	0.72908	0.73567

Ablation Study

Model	F2-Score (Public)
MTM (remove diversity)	0.77197
MTM (remove CPC)	0.77362
MTM (remove masking strategy)	0.77411
MTM (remove AC)	0.77866
MTM (remove CC)	0.77889
MTM (remove LP)	0.78007
MTM	**0.78255**

Methodology

- x : input images
- y : input labels
- g : output of ResNet18
- h : output of Bi-GRU
- $\hat{y}$: output of classifier
- res, rnn, cls : function of ResNet18, bi-GRU, and classifier.

Tasks:

1. **Multi-label Classification:** Asy is the asymmetric loss function.

$$L_1 = Asy(y, \hat{y})$$

$$Asy(y, \hat{y}) = -y(1-\hat{y})^{\gamma_+} \log(\hat{y}) - (1-y)\hat{y}^{\gamma_-} \log(1-\hat{y})$$

2. **Auxiliary Classification (AC):** Remove the RNN layer and only use g to predict labels.

$$L_2 = Asy(y, cls(g))$$

3. **Number of Label Prediction (LP):** Predict the number of labels for each image. The objective function is cross entropy loss.

$$L_3 = CE(y, cls_2(g))$$

4. **Class-Correlation Classification (CC):** It uses the Euclidean distance between g and label embedding to predict the labels. The distance between a feature g and k-th label embedding should be small if the feature contains k-th label. (Asymmetric Loss)

$$d_k = f_d(g, c_k)$$
$$d = [d_1, d_2, \dots, d_K] \text{ (K=5)}$$
$$q = -\text{softmax}(d)$$
$$L_4 = Asy(y, q)$$

5. **Contrastive Predictive Coding (CPC):** This is an unsupervised learning task which is used to extract useful representations from high-dimensional data. This task aims to predict $g_{t+1}, g_{t+2}, \dots, g_{t+T}$ from h_t (InfoNCE Loss)

$$L_5 = -E\left[log \frac{f_k(g_{t+k}, h_t)}{\sum_{g_j} f_k(g_j, h_t)}\right]$$

Loss functions:

- $L = \alpha L_1 + \beta L_2 + \gamma L_3 + \mu L_4 + \nu L_5$
- $\alpha=\beta=1, \gamma=\mu=\nu=0$ (Pretraining)
- $\alpha=\beta=\gamma=\mu=\nu=1$ (Training MTM)

Initialization:

- **Pretrained model using MC and AC tasks only (L1 and L2)**

Generalization:

- **Masking :**
  Choose p_1% of the image positions randomly. If i-th image is chosen, we replace g_i with all zeros.
- **Diversity :**
  Sample a pair of images (z_1,z_2) from training data with following conditions.
  - *i.* z_1 and z_2 are adjacent to each other, with z_2 behind z_1
  - *ii.* z_2 must always have one more label than z_1
  - *iii.* z_1 and z_2's label sequences must be identical except for the one label z_2 has that z_1 does not. (z_2 contains an additional label l which does not exist in z_1 , [0,1,1,0,0] and [0,1,1,0,1])

Therefore, $\boldsymbol{info(l) = z_2 - z_1}$

Procedures for constructing a new image sequence with image sequence $\{x_1, x_2, \dots, x_m\}$ and labels $\{y_1, y_2, \dots, y_m\}$:

- i. Let n = m * p_2
- ii. Choose n images from the sequence randomly.
- iii. Sample n pairs of images which satisfy the conditions above. Calculate $info(l_k)$ for each pairs of images and obtain $\{info(l_{k_1})_1, info(l_{k_2})_2, \dots, info(l_{k_n})_n\}, k_1, k_2, \dots, k_n \in \{1, \dots, K\}$
- iv. Add $info(l_{k_i})_i$ to its corresponding i-th chosen images in step ii.
- v. Add l_{k_i} to the labels of i-th chosen images if it doesn't exist already

Similar to scheduled sampling strategy in NLP, for each patient, we define a probability ϵ to decide whether the model is trained by the original image sequence or the generated image sequence (by applying masking or adding additional labels as described above)