B-2-3

空間的特徴に着目した知識逆蒸留による嚥下機能評価の検討

鈴木 晴仁（萩原研究室）

1　はじめに

近年，高齢化による患者の増加や医師不足に伴い，AIを用いた医用画像処理が注目されている．特に，耳鼻咽喉科では嚥下障害患者の増加が著しく，臨床リハビリテーションにおける食塊検知モデルの需要が高まっている．

摂食嚥下リハビリテーションでは，嚥下機能を評価するにあたり咽頭残留を検出することが重要である．嚥下機能評価のゴールドスタンダードは嚥下造影検査（VideoFluoroscopic examination of swallowing: VF）であるが，設備や侵襲性，造影剤誤嚥等の問題から，日常的な使用は制限されている．また，医師による診断はVFに基づく定性的評価が中心であり，嚥下機能を客観的に評価するための定量的指標は未だ確立されていない．

本研究では，空間的な選択性を付与する注意機構を導入した知識逆蒸留に基づく嚥下機能評価法を提案する．本手法は，VFに代わる非侵襲的かつ定量的な嚥下機能の評価手法である超音波検査の枠組みに導入することが可能な教師なし食塊検知モデルである．実験では，嚥下超音波画像データセット [1] における検知精度および領域一致度を評価することで，提案手法の有効性を示す．

2　提案手法

本研究は，先行研究 [2] に着想を得て，One-Class Classification（OCC）に基づく食塊検知手法を提案する．OCC設定に従うモデルは，食塊を含まない食道状態を正常と定義し，その特徴分布を学習する．推論時には，健康状態を基準とする医師の診断フローを模倣し，学習した正常特徴分布から逸脱する入力を異常（食塊）として検知する．以下では，提案手法が採用する知識逆蒸留の概要ならびに導入する注意機構 Convolutional Block Attention Module（CBAM）[3] について述べる．

知識逆蒸留： 知識逆蒸留では，従来の知識蒸留における教師-生徒（T-S）モデルにエンコーダ-デコーダ構造を採用する．教師は豊富な特徴表現を抽出することを目的とする．生徒は教師の対称的かつ反転した構造を持ち，教師の出力を入力として教師の挙動を模倣することを目指す．推論時にはマルチスケール特徴に基づく蒸留を用いることで，局所的かつ領域的な異常を示す．

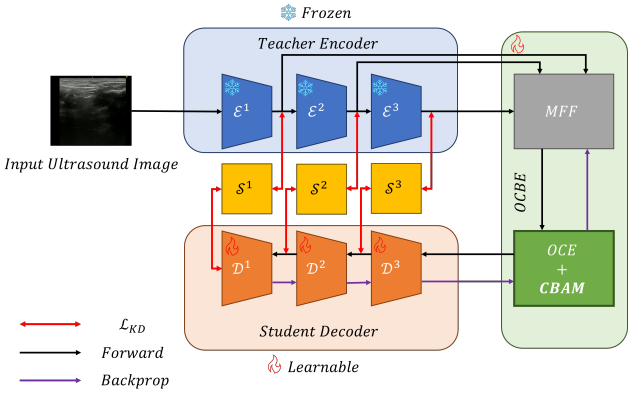

図 1: 提案手法の概要．

One-Class Bottleneck Embedding（OCBE）： OCBE モジュールは，教師が抽出したマルチスケール特徴を集約する Multi-Scale Feature Fusion（MFF）ブロックと，MFF ブロックの出力を圧縮する One-Class Embedding（OCE）ブロックで構成される．多くのノイズを含む超音波画像に対し，本モジュールは正常サンプルのみを用いた学習により情報ボトルネックとして機能し，正常特徴を効果的に保持することが可能である．

知識蒸留損失： OCBE モジュールと生徒デコーダは知識蒸留損失により共同で最適化される．まず，T-S モデル間のチャネル軸に沿ったベクトル単位のコサイン類似度損失を計算し，異常度マップ $\mathbf{M^k}\in\mathbb{R}^{H_k\times W_k}$ を得る．マルチスケール蒸留を考慮したスカラー損失関数 $\mathcal{L}_{KD}$ は，全スケールの異常度マップを加算することで定義される:

$$\mathcal{L}_{KD}=\sum_{k=1}^{K}\left\{\frac{1}{H_kW_k}\sum_{h=1}^{H_k}\sum_{w=1}^{W_k}M^k(h,w)\right\}. \qquad (1)$$

K は考慮する特徴階層数を表す．画像全体の異常度マップは，$\mathbf{M^k}$ を画像サイズにアップサンプリングした後，ピクセルレベルで全スケールを加算して得る．

CBAM: CBAM は，チャネルアテンションモジュールと空間アテンションモジュールの逐次的配置からなる．中間特徴マップ $\mathbf{F}\in\mathbb{R}^{C\times H\times W}$ に対して，チャネルアテンションマップ $\mathbf{M_c}(\mathbf{F})\in\mathbb{R}^{C\times 1\times 1}$ と空間アテンションマップ $\mathbf{M_s}(\mathbf{F}')\in\mathbb{R}^{1\times H\times W}$ を順次推定し，出力 $\mathbf{F}''$ を得る:

$$\mathbf{F}'=\mathbf{M_c}(\mathbf{F})\otimes\mathbf{F},\ \mathbf{F}''=\mathbf{M_s}(\mathbf{F}')\otimes\mathbf{F}'. \qquad (2)$$

$\otimes$ は要素ごとの積を表し，$\mathbf{M_c}(\mathbf{F})$ は空間次元，$\mathbf{M_s}(\mathbf{F}')$ はチャネル次元に沿ってブロードキャストされる．

表 1: 嚥下超音波画像データセット [1] における食塊検知精度の定量的評価（AUROC（%）↑ / PRO（%）↑）.

Category	Bread	Cracker	Jelly	Pudding	Soda	Yogurt	Yokan	Avg
AD [2]	83.6 / 62.2	81.7 / 53.9	82.6 / 57.3	83.0 / 56.0	78.3 / 41.7	**85.1** / **60.6**	**84.7** / **59.2**	82.7 / 55.8
Ours	**86.2** / **64.4**	**84.8** / **55.4**	**85.3** / **59.9**	**85.4** / **59.5**	**81.2** / **44.7**	84.8 / 59.4	84.5 / 59.0	**84.6** / **57.5**

Channel Attention Module: チャネルアテンションは，入力に対して注目すべき特徴表現に焦点を当てる．特徴マップ $\mathbf{F}$ に対して平均／最大プーリングをそれぞれ適用後，MLP を通し，要素ごとの和で統合する:

$$\begin{aligned}\mathbf{M_c}(\mathbf{F}) &= \sigma\left(\mathrm{MLP}(\mathrm{AvgPool}(\mathbf{F})) + \mathrm{MLP}(\mathrm{MaxPool}(\mathbf{F}))\right) \\ &= \sigma\left(\mathbf{W_1}(\mathbf{W_0}(\mathbf{F^c_{avg}})) + \mathbf{W_1}(\mathbf{W_0}(\mathbf{F^c_{max}}))\right)\end{aligned} \tag{3}$$

σ は活性化関数，$\mathbf{W_0} \in \mathbb{R}^{C/r \times C}, \mathbf{W_1} \in \mathbb{R}^{C \times C/r}$ は MLP における特徴変換行列，r（$= 16$）は次元縮小率を表す．

Spatial Attention Module: 空間アテンションは，チャネルアテンションで失われる空間情報を補完する．特徴マップ $\mathbf{F}'$ に対してチャネル方向の平均／最大プーリングをそれぞれ適用後，連結し，畳み込み層を通す:

$$\begin{aligned}\mathbf{M_s}(\mathbf{F}') &= \sigma\Big(f^{7\times7}\big([\,\mathrm{AvgPool}(\mathbf{F}');\ \mathrm{MaxPool}(\mathbf{F}')\,]\big)\Big) \\ &= \sigma\Big(f^{7\times7}\big([\,\mathbf{F'^s_{avg}};\ \mathbf{F'^s_{max}}\,]\big)\Big)\end{aligned} \tag{4}$$

$f^{7\times7}$ はカーネルサイズ 7×7 の畳み込み演算を表す．

CBAM を用いた知識逆蒸留に基づく T-S モデル: 提案手法のモデルは，図 1 に示すように，学習時にパラメータを固定した教師エンコーダと，CBAM を導入した学習可能な OCBE モジュールおよび生徒デコーダで構成される．CBAM を組み込んだ OCBE モジュールは，教師エンコーダが抽出するマルチスケール特徴を集約，圧縮，精緻化し，正常特徴を効果的に保持した潜在表現を生成する．生徒デコーダは OCBE モジュールの出力から教師エンコーダの出力を再構成するように学習を行う．教師が食塊特徴を正確に抽出する一方，OCBE モジュールおよび生徒は正常サンプルにより最適化されているため，推論時に異常サンプルが入力された場合，T-S モデル間の再構成誤差を食塊として検知可能である．

3 評価実験

実験設定: 本研究で使用する嚥下超音波画像データセット [1] は，7 種類の食塊を含む食道および空の食道における嚥下反射を捉えた超音波画像 2395 枚（学習用 652 枚，テスト用 1743 枚）で構成される．評価指標には，画素単位の検知精度／領域一致度を測る Pixel AUROC／AUPRO を用いる．

実験結果: 表 1 に定量的評価，図 2 に定性的評価を示す．提案手法は，多くのカテゴリで競合手法 [2] を上回ることが確認できる．CBAM を組み込んだ OCBE モジュールはチャネルおよび空間方向の注意を与え，正常特徴を強調しノイズ的成分を相対的に弱める．これにより，図 2 に示すように，正常特徴への選択性が安定し精緻化された出力を得る．また，推論時にノイズや異常特徴が CBAM に入力されると不適切な強調を生じる可能性があるが，実際にはこの懸念よりも正常特徴に対する強調効果が優勢となり，結果的に検知精度の向上に寄与していると考えられる．

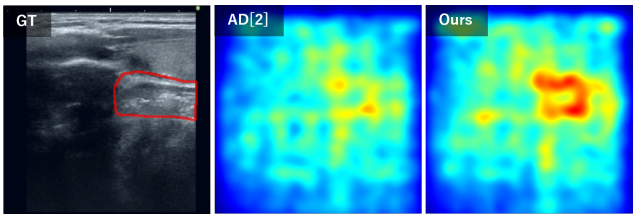

図 2: 競合手法 [2] ならびに提案手法の定性的評価．左から GT，競合手法，提案手法の異常度マップを示す．

4 まとめ

本研究では，注意機構 CBAM を導入した知識逆蒸留に基づく教師なし食塊検知モデルを提案した．評価実験の結果，提案手法の空間的な選択性を付与するモデル構成は，多くのカテゴリで検知精度の改善に貢献することが確認された．今後は，本手法を摂食嚥下リハビリテーションの臨床応用へ展開することを念頭に，実環境における性能評価ならびにモデルの高精度化を目指す．

参考文献

[1] Q. Gao, Y. Hagihara, M. Sasaki, and K. Hotta, “Attention-guided food bolus segmentation in ultrasound imaging for dysphagia rehabilitation,” IEEJ Transactions on Electrical and Electronic Engineering, vol.20, no.11, pp.1757–1765, 2025.

[2] Q. Gao, Y. Hagihara, M. Sasaki, C. Gu, and K. Hotta, “Unsupervised food bolus detection for ultrasound images via reverse distillation,” Proceedings of the GCCE, IEEE, IEEE, sep 2025.

[3] S. Woo, J. Park, J.Y. Lee, and I.S. Kweon, “Cbam: Convolutional block attention module,” Proceedings of the ECCV, pp.3–19, 2018.