表画像の補正と大規模言語モデルによる表構造解析手法の改良

納田　朋享† 金澤　輝一†† 上野　史††† 太田　学†††

† 岡山大学大学院環境生命自然科学研究科　〒 700–8530 岡山県岡山市北区津島中 3–1–1
†† 国立情報学研究所コンテンツ科学研究系　〒 101–8430 東京都千代田区一ツ橋 2–1–2
††† 岡山大学学術研究院環境生命自然科学学域　〒 700–8530 岡山県岡山市北区津島中 3–1–1
E-mail: †pwnk12a9@s.okayama-u.ac.jp, ††tkana@nii.ac.jp, †††{uwano, ohta}@okayama-u.ac.jp

あらまし　実環境で撮影された表画像には，撮影角度に起因する幾何学的な歪みが生じやすく，これは表構造解析の精度低下を招く要因となる．そこで本研究では，実環境での撮影を模して人工的に台形歪みを加えた表画像を生成し，生成した表画像の歪みを補正する手法を提案する．また，近年様々な分野で活用されている大規模言語モデル（LLM）に着目し，Ye らが提案した表構造解析モデルである TableMASTER の表構造解析結果を，LLM を用いて修正する手法を提案する．評価実験では，ICDAR 2021 Competition on Scientific Literature Parsing（ICDAR2021-SLP）のテストデータ 1,000 件の表画像を解析し，表構造解析精度を Tree-Edit-Distance-based Similarity（TEDS）を用いて評価した．実験の結果，台形歪みを加えたテストデータを用いた検証では，台形歪みの補正によって TEDS の値が 0.6333 から 0.7205 へ，画像の構造的類似度を示す Multi-Scale Structural Similarity Index（MS-SSIM）は元の表画像との比較において 0.3272 から 0.5648 へ改善した．また，歪みのない表画像のテストデータから得られた解析結果に対する検証では，LLM による修正によって TEDS の値が 0.9596 から 0.9599 へ向上した．

キーワード　表構造解析，表画像，大規模言語モデル（LLM）

1 はじめに

学術論文において，実験結果や統計情報は表としてまとめられることが多い．表構造を解析できれば，表から視覚的に優れたグラフへの自動変換 [2] や情報の抽出 [3]，複数の表の集約といった応用が可能となり，論文理解の効率化に大きく寄与する．しかし，表の様式は一般に著者によって異なり多様であるため，罫線の有無やマルチカラムセルなどを考慮した表構造解析が必要となる．また，表画像を解析できれば，スクリーンショットや古い文献のスキャン画像，手書きの表など，メタデータを持たない表に対しても構造解析が可能となり，その汎用性は高い．このような背景から，表画像を入力とする表構造解析の研究が活発に行われている [4] [5].

しかし，実環境での利活用を想定した場合，入力される表画像は真正面から影や歪みなく撮影された理想的な画像であるとは限らない．撮影時のカメラ角度に起因する台形歪みなどが生じている場合，既存のモデルでは正しく構造を認識できないことが多い [6].

近年，大規模言語モデル（Large Language Model，LLM）およびマルチモーダル LLM の急速な発展により，大規模事前学習で獲得された高度な言語知識と推論能力を活用した文書理解が可能となっている．これにより，従来の OCR やルールベース手法では困難であった文書の意味的理解が実現されつつある．具体的には，文脈や意味整合性を考慮した OCR 結果の修正 [7]，表やチャートからの構造化データの復元 [8]，さらに文書質問応答タスクにおいても，視覚情報と数値情報を統合的に理解する能力の有効性が示されている．これらの性能は，DocVQA [9] などの評価基盤を通じて広く検証・活用されている．これらの研究成果は，従来の解析モデルによって得られた出力結果に対し，LLM を用いて意味的な検証および修正を行う後処理が，文書および表情報の実用的な利活用において重要な役割を果たすことを示唆している．

筆者らは先行研究において，Ye らの表構造解析手法 [4] の分析および追加学習による改良について報告した [1]．そこで本研究では，Ye らの表構造解析手法 [4] を基盤モデルとして採用し，実環境での頑健性と解析精度の向上を目的として，台形歪みの補正および LLM による表構造情報の修正を提案する．提案手法は，台形歪みを加えた表画像の補正と LLM を用いた表構造解析結果の修正の 2 つから構成される．

台形歪みの補正では，人工的に台形歪みを付与した画像データセットを生成し，それを用いて画像を正規化する．本処理により，歪みによって損なわれた画像の幾何学的特徴を復元し，表構造解析モデルが想定する入力品質へ近づける．LLM を用いた修正では，表構造解析モデルが出力した HTML コードを入力とし，LLM の推論能力を活用して表構造の誤りおよびセルテキストの誤りを修正する．

評価実験では，ICDAR 2021 Competition on Scientific Literature Parsing（ICDAR2021-SLP）[10] のテストデータを解析する．評価指標は 2 つある．まず，台形歪みの補正では Multi-Scale Structural Similarity Index（MS-SSIM）を用いて画像の補正品質を評価する．Tree-Edit-Distance-based Similarity（TEDS）を用いて表構造解析精度を評価する．LLM による表構造解析結果の修正実験では，Tree-Edit-Distance-based Similarity（TEDS）を用いて表構造解析精度を評価する．

本稿の構成は以下の通りである．第 2 節では関連研究につい

て述べる．第 3 節では台形歪みを加えた表画像の補正手法，第 4 節では LLM を用いた表構造解析結果の修正手法について述べ，第 5 節では提案手法の有効性を検証するための評価実験について説明する．第 6 節でまとめる．

2 関 連 研 究

2.1 表画像を入力とする表構造解析

Ye らは，文書解析タスクである ICDAR2021-SLP において，表画像の表構造解析を「テキスト行検出」,「テキスト行認識」，および「セルへのテキスト割り当て」という 3 つのサブタスクに分割して処理するモデルである TableMASTER を提案した [4]．この手法では，表構造認識およびテキスト行認識のモデルとして，高精度な画像テキスト認識モデルである MASTER [11] を改良したモデルを採用している．特に表構造認識においては，HTML タグの予測とバウンディングボックスの予測を並列に行う構造を導入した．また，テキスト行検出には，任意の形状のテキストや近接するテキスト行を効果的に識別可能な PSENet [12] を利用した．最終的な HTML 生成に向けたボックス割り当てフェーズでは，検出されたテキストボックスとセルを関連付けるために，まず中心点ルール，次に IoU ルール，最後に距離ルールを順次適用するという，3 段階の階層的なマッチング規則を採用した．ICDAR2021-SLP [10] のテストデータセットを用いた評価実験において，提案手法は TEDS の値が 0.9632 となった．

Smock らは，物体検出手法である Detection Transformer（DETR）[13] を表構造解析問題に応用した Table Transformer を提案した [14]．Table Transformer は，表内の行，列，および見出し領域をそれぞれ独立した検出対象として直接推定し，それらの幾何的配置関係に基づいて表の論理構造を復元する手法である．彼らは，従来の表構造解析用データにおいて，行・列・セル間の対応関係に不整合が存在することが認識精度の低下を招いていると指摘し，これらの対応関係が一貫して定義された大規模表画像データセットである PubTables-1M を新たに構築し，これを用いて学習した．この枠組みにより，複数の行や列にまたがるセルを含む表や，罫線を持たない表に対しても，罫線情報に依存することなく，検出された行と列の交差関係から表構造を安定して推定できることを示した．実験の結果，Table Transformer は PubTabNet を用いた表構造解析において TEDS の値が 0.9360 となった．

2.2 歪みのある画像の補正

Bandyopadhyay らは，文書画像の幾何学的歪み補正において，畳み込みニューラルネットワークの一種である U-Net を拡張した RectiNet を提案した [15]．RectiNet では，画像内のエッジや境界線の詳細を捉えるための Gated Network と，密な歪み補正グリッドを予測する際にチャンネル間の情報の混在を防ぐための分岐型 U-Net を導入している点が特徴である．RectiNet は，約 8,000 枚の合成データによる学習でありながら，DocUNet データセット [16] を用いた評価において多重解像度での構造的類似度を測る Multi-Scale Structural Similarity Index（MS-SSIM）や局所的な歪みを測る Local Distortion（LD）といった指標で最先端の性能を達成した．

Zhu らは，歪んだ表画像では表構造解析の精度が低下するという課題に対し，表の構造的特徴を活用した U-Net ベースの新しい歪み補正モデルを提案した [17]．Zhu らは，まず，モデルが表の構造に着目できるよう，セルや表の罫線といったキー要素をセグメンテーションするモジュールを導入した．また，表の線分性を保つためには局所領域ではなく画像全体の歪みを把握する必要があるため，エンコーダに Transformer を組み込み，全体的な歪みを捉える能力を強化した．さらに，歪み補正の過程で生じるぼやけが可読性や評価指標に悪影響を与える点に着目し，軽量な鮮鋭化を後処理として適用することで最終的な画質向上を実現している．加えて，表画像の歪み補正に特化したデータセットが存在しなかったため，PubTabNet の HTML を再レンダリングし，歪みを加えることで 12,000 枚の合成データセットを独自に構築した．実験の結果，提案手法は従来の文書補正手法よりもすべての画質評価指標で優れた性能を示し，特に MS-SSIM では約 15 ポイントの大幅な改善が得られた．さらに，補正後の画像を用いて表構造を解析すると，TEDS スコアが約 6 ポイント向上した．

2.3 大規模言語モデルを用いた表構造解析結果の修正

Ren らは，表構造解析モデルが出力した結果を後処理によって修正，改善するアプローチとして，TableGLM [18] を提案した．Ren らの手法は，まず Transformer ベースの表構造解析モデルを用いて表画像から表の HTML コードを生成する．次に，この HTML コードを，表構造とテキスト内容の修正タスクに特化させてファインチューニングした LLM である TableGLM に入力する．TableGLM は，ChatGLM3-6B モデルを基盤とし，表構造解析モデルが生成した HTML コードと，それに対応する正解の HTML コードをペアにしたデータセットを用いて学習している．実験により，TableGLM による修正ステップを加えることで，PubTabNet データセットにおいて TEDS の値が平均 3.1 ポイント向上した．

Zhang らは，表構造解析を含む多様な表関連タスクに対応するための汎用モデルとして，TableLlama [19] を提案した．Zhang らの手法は，Llama 2 モデルを基盤とし，表構造の生成や修正を含む 15 種類の表タスクに対して適切に応答できるように追加学習を行う指示チューニングを行っている．TableLlama は，260 万件以上の表画像とテキストのペアを含む TableInstruct データセットを用いて学習されており，これには PubTabNet などの主要な表データセットから構築された構造解析タスクが含まれている．実験により，TableLlama は PubTabNet を含む複数のベンチマークにおいて，GPT-3.5 や GPT-4 などの汎用 LLM と同等以上の表構造理解能力を示し，7B パラメータという軽量なモデルでありながら高い TEDS の値を達成した．

3 台形歪みを加えた表画像の生成および補正

3.1 表の構成要素と表構造

本稿で扱う表画像は HTML タグによって表構造が定められる．本稿で扱う表および表に対応する HTML コードの例を図 1 に示す．

HTML 形式の表はヘッダ部分とボディ部分からなり，`<thead>...</thead>`は表の列の見出しを表すヘッダ行の部分を表しており，`<tbody>...</tbody>`は表のその下にあるボディ部分を表している．なお，他にも表のフッタ部分を表す`<tfoot>...</tfoot>`もあるが，本稿では扱わない．以下に本稿で扱う HTML タグをまとめる．

- `<thead>...</thead>`：表のヘッダ部分
- `<tbody>...</tbody>`：表のボディ部分
- `<b>...</b>`：太字表記
- `<i>...</i>`：斜体表記
- `<sup>...</sup>`：上付き文字
- `<sep>...</sep>`：行区切り文字
- `<tr>...</tr>`：表の行
- `<td>...</td>`：表のデータの各要素を表すセル
- `<td rowspan="n">...</td>`：n 個の垂直方向の結合セル
- `<td colspan="n">...</td>`：n 個の水平方向の結合セル

3.2 台形歪みを加えた表画像の生成

実環境におけるカメラ撮影では，撮影角度や遠近法の影響により，表画像が台形状に歪むことが多い．このような台形歪みは，画像中の水平性や垂直性が保持されていることを前提とする既存の表構造解析モデルにとって障害となり，解析精度の著しい低下を招く要因となる．そこで，実環境下での撮影条件を考慮した頑健な表構造解析の実現を目的として，台形歪みを加えた表画像を人工的に生成し，その補正を試みる．

本研究では，入力された表画像に対して射影変換行列を適用することで，台形歪みを有する表画像を生成する．具体的には，入力画像の四隅を変換前の基準点とし，各頂点を画像の幅および高さに対する一定割合の範囲内でランダムに変位させた座標を変換後の対応点として設定する．これらの対応点に基づいて射影変換行列を算出し，画像全体に適用することで，台形歪みを加えた表画像を生成する．

また，変換後の画像において画素が画像領域外へはみ出すことを防ぐため，変換後の四隅座標の最小値および最大値を用いて出力画像サイズを決定し，座標系を平行移動によって正規化した．この処理により，台形歪みを加えた後も表画像全体が出力画像内に収まるようにしている．以上の処理により，入力された表画像から台形歪みを加えた表画像を生成した．

歪ませた表画像の例を図 2 に示す．図 2（a）から（b）が生成される．

3.3 台形歪みを加えた表画像の補正

本節では，撮影条件や視点の影響により台形歪みが加わった表画像に対し，表構造解析の前処理として，表領域を幾何学的に正規化する補正手法を提案する．台形歪みを加えた表画像の補正処理の概要を図 3 に示す．なお，本節で述べる処理はすべて OpenCV を用いて実装した．

提案手法では，まず入力画像に対して表の罫線を強調するための前処理を行う．具体的には，局所的な輝度分布に基づいて二値化閾値を動的に決定する手法である適応的二値化を適用し，濃淡変化の影響を抑えつつ罫線を抽出する．続いて，膨張および収縮といった形態学的処理を適用し，擦れやノイズによって分断された線分の接続を促進する．次に，前処理後の画像に対して水平線分を検出する．提案手法では，画素単位よりも細かい精度で線分区間を検出できる手法である Line Segment Detector（LSD）を適用し，表罫線を高精度に検出する．しかし，ノイズの影響により，LSD では十分な線分が得られない場合がある．そのような場合には，確率的に線分を探索する手法である確率的 Hough 変換を用い，検出精度の低下を補完することで，線分検出の失敗を抑制する．検出された線分群のうち，画像内でほぼ水平方向に延びている成分のみを選択し，台形歪みに対して安定な水平線分を抽出する．抽出された水平線

(a) 表

野菜の種類	商品情報	
	個数	値段
人参	5	200
ミニトマト	20	400

(b) 表に対応する HTML コード

```
<table><thead>
<tr><td rowspan="2"><b>野菜の種類</b></td>
<td colspan="2"><b>商品情報</b></td></tr>
<tr><td><b>個数</b></td><td><b>値段</b></td></tr>
</thead><tbody>
<tr><td>人参</td><td><i>5</i></td><td>200</td></tr>
<tr><td>ミニトマト</td><td>20</td><td>400</td></tr>
</tbody></table>
```

図 1: 本稿で扱う表とその HTML コードの例

(a) 元の表画像

Parameter	Value
Self-inductance	$L = 6$ [mH]
Rated current	$I = 8$ [A] (AC, 50/60 Hz, sinus wave)
Current density	$J < 3$ [A/mm^2]
Ambient temperature	$\vartheta_a = +40$ [°C]
Self-resonance frequency range	$f_{crit} = 80 \div 140$ [kHz]

(b) 台形歪みを加えた表画像

Parameter	Value
Self-inductance	$L = 6$ [mH]
Rated current	$I = 8$ [A] (AC, 50/60 Hz, sinus wave)
Current density	$J < 3$ [A/mm^2]
Ambient temperature	$\vartheta_a = +40$ [°C]
Self-resonance frequency range	$f_{crit} = 80 \div 140$ [kHz]

図 2: 生成した台形歪みを加えた表画像の例（表画像の出典：PubTabNet）

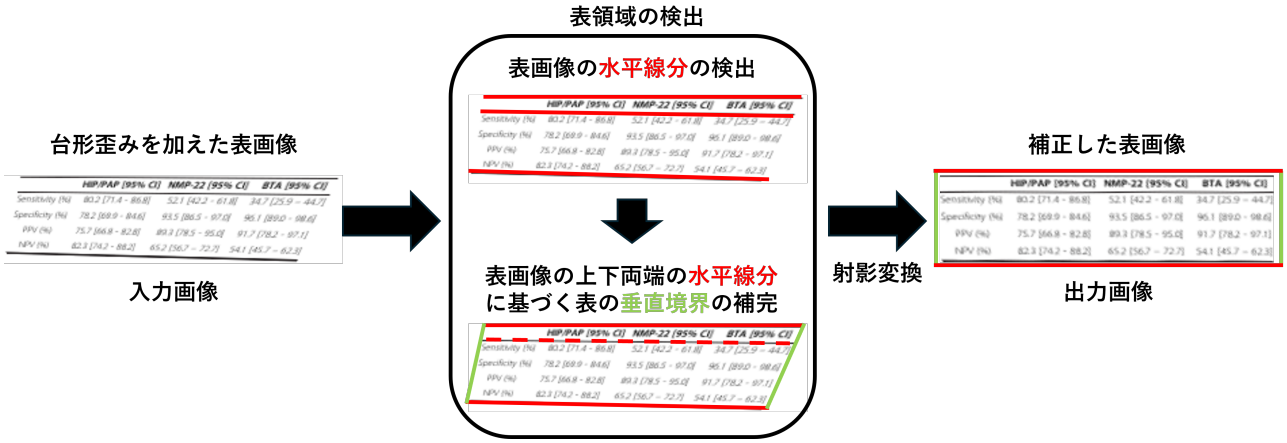


図 3: 台形歪みを加えた表画像の補正手法の概要 (表画像の出典：PubTabNet)

分は，罫線の欠損や点線化の影響により，複数の短い線分として検出される場合がある．そこで，各線分の位置および傾きの類似性に基づいて線分を統合し，同一の線分要素として扱う．続いて，各線分群の端点を点群として扱い，点群の分布特性に基づいて近似線分を推定する手法である主成分分析を適用することで，各クラスタを代表する近似線分を算出する．算出された線分群の中から，画像座標系において最上部および最下部に位置する線分をそれぞれ表領域の上端および下端と定義する．続いて，表の左右端に位置する縦方向の罫線を直接検出するのではなく，特定された上端および下端の水平線分の始点と終点を対応付け，それらを結ぶことで左右の境界を補完的に決定する．最後に，対応する 4 点から射影変換行列を算出する．得られた射影変換行列を用いて画像を幾何変換することで，台形歪みを補正した表画像を生成する．

提案手法は水平線分の検出と統合に重点を置くため，一般的な台形補正とは異なり，表の左右端に位置する縦方向の罫線が存在しない表や，罫線が部分的に欠損している表に対しても，水平方向の情報のみから頑健に表領域を推定することが可能である．

4 LLM を用いた表構造解析結果の修正

4.1 提案する修正手法の概要

本節では，表構造解析モデルが出力した HTML 形式の表構造解析結果に対し，外部の参考情報を活用した LLM による修正手法を提案する．LLM を用いた表構造解析結果の修正の概要を図 4 に示す．本稿では，Ye らによって提案された表構造解析モデルである TableMASTER [4] によって出力された表構造解析結果の HTML コードを LLM への入力として，専門用語集および過去の解析の事例を外部知識として LLM に与えることで，修正した HTML コードを出力する．その後，出力された修正案と入力 HTML コードとの幾何学的整合性や内容の一貫性に基づく比較により，修正の安定性を担保する．この修正の目的は，OCR 由来の軽微な単語誤りの訂正と一部のセルの配置ミスの改善である．

4.2 LLM に与える外部知識

提案手法では，LLM が適切な修正を行うための外部知識として，2 種類の情報を用いる．1 つ目は，PubTabNet および ICDAR2021-SLP [10] のテストデータセットに関連する生理学分野の専門用語集である Medical Subject Headings（MeSH）である．2 つ目は，TableMASTER による PubTabNet の検証データの予測結果とその正解データの HTML コードの組の情報である．

MeSH は，米国国立医学図書館が作成・管理するシソーラスであり，医学用語が階層的に定義されている．PubTabNet および ICDAR2021-SLP [10] のテストデータセットに含まれる表データは医学論文由来であるため，セル内の記述には専門的な薬剤名，疾患名，解剖学用語などが頻出する．MeSH を与えることで，OCR の誤りによる専門用語のスペルミスを LLM が検知し，正しく修正できるようになることを期待する．

TableMASTER による予測結果とその正解データの HTML コードの組は，PubTabNet の検証データセットの表 9,116 件から抽出したものである．この情報を与える目的は，表構造解析モデルが犯しやすい誤りのパターンを LLM に具体例として示すことで，入力された HTML コードに対する適切な修正を促すことである．

4.3 RAG による表構造解析結果の修正手法

4.2 節の外部知識を本稿では Retrieval Augmented Generation（RAG）[20] を利用して LLM に与える．RAG は，外部の文書集合から関連情報を検索し，その結果を LLM の入力として与える枠組みである．RAG 手法は，検索と生成を統合的に学習する end-to-end 型と，検索結果をそのままプロンプトに付与するコンテキスト注入型に大別される．前者は高い性能を示す一方で再学習が必要となる．そこで提案手法では，既存の LLM を変更することなく外部知識を導入できるコンテキスト注入型を採用する．これにより，専門用語や誤り訂正の事例を動的にプロンプトへ組み込むことが可能となる．

修正処理の具体的な手順は以下の通りである．まず，4.2 節で述べた外部知識を事前に整備する．TableMASTER による予測結果と正解データの HTML コードの組については，入力

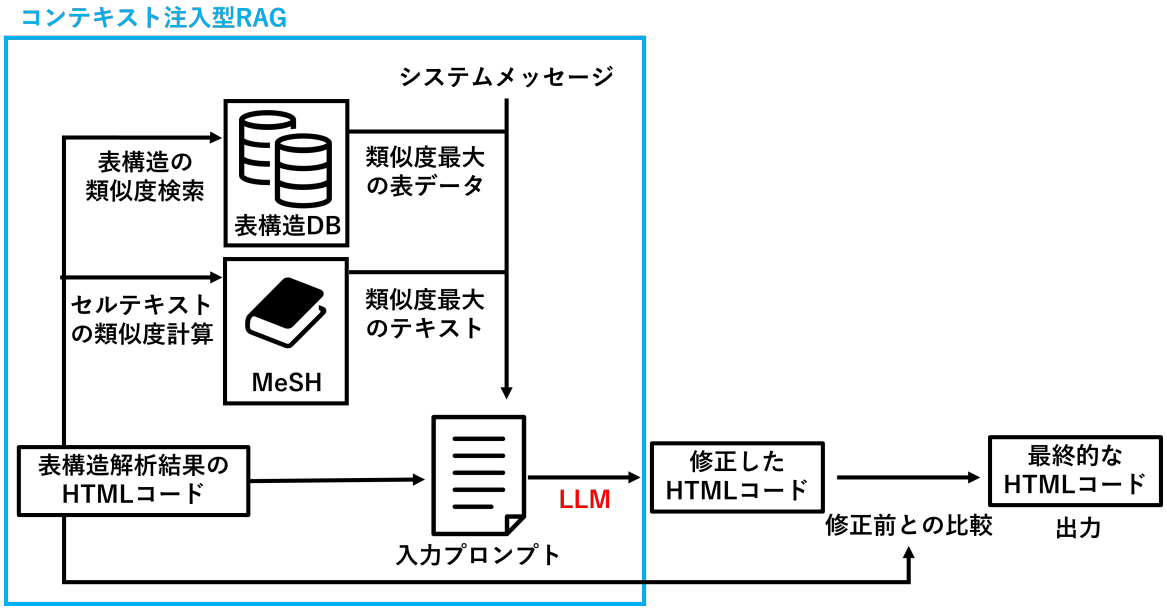


図 4: LLM を用いた表構造解析結果の修正手法の概要

HTML コードと表構造が似ている表を外部知識とするため，セル内のテキストを除去し，表構造を表すタグのみからなる HTML コードを生成する．また，MeSH については，各専門用語を収集する．これらの表構造タグのみの HTML コードおよび MeSH の各専門用語は，Sentence Transformers の文埋め込みモデルを用いてベクトル化され，データベース (DB) に格納される．

次に，修正対象となる表の HTML コードが入力されると，まず入力 HTML コードからセル内テキストを除去し，表構造を表すタグのみからなる HTML コードを生成する．この表構造のみの HTML コードをクエリとして，データベース（DB）に格納された正解の表構造 HTML ベクトルとのコサイン類似度を計算し，最も類似度の高いものを取得する．その結果として，表構造が最も類似する表の HTML コードの予測結果とその正解データの HTML コード，および行単位の類似した表の予測結果と正解の HTML コードの差分を参照情報として抽出し，LLM への入力プロンプトに付与する．一方，MeSH に基づく専門用語の参照情報については，入力 HTML コードから抽出したセルテキストをクエリとして，各専門用語とのコサイン類似度を計算し，類似度が最大となる用語を取得する．取得した用語情報は，OCR 由来の専門用語の誤り訂正を促すための参照情報として，表構造に関する参照情報とともに LLM への入力プロンプトに付与する．

その後，LLM への入力プロンプトを生成する．プロンプトは，LLM の役割と修正条件を定義したシステムメッセージ，外部知識，および修正対象の HTML コードから構成される．そして，LLM は生成されたプロンプトに基づき，外部知識を手掛かりとして入力 HTML コードを解析し，表構造に含まれる誤りを必要に応じて修正することで，構造的整合性が保たれた修正した HTML コードを出力する．

ただし，大規模言語モデル（LLM）による生成結果は常に構造的に正しいとは限らず，過度な修正による事実に基づかない出力や構成要素の不整合が発生する可能性がある． そこで，修正結果と TableMASTER による表構造の解析結果を照らし合わせ，形状の正しさや内容の一致具合に基づいて検証し，最終的な HTML 記述を選択する後処理を行う．具体的には，文字列の類似性を評価する difflib を用いた指標や行数の変動，重要な見出し情報の維持状況を確認し，内容の改ざんを防ぐ安全策を設けている．その上で，各行の横幅のばらつきを数値化した指標や空行の有無を算出し，表としての幾何学的な整合性が向上したかを定量的に評価する．この処理により，有効な修正案がない場合や，格子の整合性が修正前よりも悪化している場合は，改悪を防ぐため修正前の入力を最終結果として採用する．これにより，モデルの不安定さを抑えつつ，構造の修復や文字の微修正など，明確な改善が見込める場合のみ修正を適用することを可能にしている．

5 評 価 実 験

5.1 評価実験の概要

3 節で提案した台形歪みの補正手法および 4 節で提案した LLM による表構造の修正手法の有効性をそれぞれ検証するために，2 つの評価実験を実施する．

1 つは，台形歪みの補正に関する評価実験である．ここでは，まず補正手法による表画像の幾何学的な復元性能を検証するために，画像そのものの品質を評価する．その上で，台形歪みを付加した表画像を入力とした場合の表構造解析精度と，表構造解析の前処理として台形歪みの補正を適用した場合の解析精度を比較することで，表構造解析における補正処理の有効性を評価する．もう 1 つは，LLM による表構造解析結果の修正効果を評価する実験である．ここでは，Ye らによって提案された表構造解析モデル [4] による解析結果の精度と 4 節の LLM に

よる修正処理を加えた場合の解析結果の精度を比較する．

5.2 実験設定

5.2.1 データセットおよび使用モデル

本実験では，表構造解析モデルとして Ye らが公開している TableMASTER [4] の学習済みモデル [1]を使用する． 同モデルは，約 50 万件の学習データから成る PubTabNet [21] を用いて学習されたものである．

評価用データセットとして，ICDAR2021-SLP [10] からランダムに抽出した 1,000 件の表画像と HTML コードのペアを用いる．台形歪みの補正に関する評価実験では，この 1,000 件の表画像に対して擬似的な台形歪みを付与した画像を生成する．そしてそれを入力として用い，提案手法による補正処理が後段の表構造解析精度に与える影響等を評価する．一方，LLM による修正効果の評価実験では，台形歪みを付与していない元の表画像 1,000 件を入力とし，TableMASTER の出力結果に対して LLM による修正処理を適用した場合の解析精度の変化を評価する．

なお，解析結果の修正に用いる LLM には Llama-3.3-70B を採用した． 推論パラメータは，temperature を 0.0，repetition penalty を 1.05，max new tokens を 8,192 に設定した．

5.2.2 評価指標

台形歪みの補正における画像の品質評価では，人間の視覚特性を考慮した画像の構造的類似度指標である Multi-Scale Structural Similarity Index（MS-SSIM）[23] を用いる．MS-SSIM は，ダウンサンプリングによって段階的に解像度を下げた複数のスケール画像を用いて画質を評価する指標である．画像 x と画像 y の間の MS-SSIM は，最大スケール M における輝度比較項 l_M と，各スケール j におけるコントラスト比較項 c_j および構造比較項 s_j を統合して，次式で定義される．

$$\text{MS-SSIM}(x,y) = l_M(x,y)^{\alpha} \prod_{j=1}^{M} c_j(x,y)^{\beta} s_j(x,y)^{\gamma} \qquad (1)$$

ここで，α, β, γ は重みパラメータであり，l, c, s は SSIM において定義される輝度，コントラスト，構造の比較関数である．

また，表構造解析の評価には，表を表す正解の HTML と予測 HTML の木構造としての類似度を測る Tree-Edit-Distance-based Similarity (TEDS) [21] を用いる．TEDS は以下の式で定義される．

$$\text{TEDS}(T_a, T_b) = 1 - \frac{\text{EditDist}(T_a, T_b)}{\max(|T_a|, |T_b|)} \qquad (2)$$

ここで，T_a は正解の表構造，T_b は予測された表構造を表す．$\text{EditDist}(T_a, T_b)$ は T_a と T_b 間の木編集距離であり，$|T_a|$ および $|T_b|$ はそれぞれの木構造におけるノード数を表す．

さらに，本実験では TEDS に加え，セルの内容を無視し，表の構造を表す HTML タグの一致度のみを評価する S-TEDS (Structural-TEDS) [22] も併せて用いる．

1：https://github.com/JiaquanYe/TableMASTER-mmocr

5.3 実験結果

5.3.1 台形歪みの補正に関する実験の結果

まず，提案手法による台形歪みの補正による画像品質の改善効果について述べる．台形歪みを加えた表画像および提案手法により補正した表画像の MS-SSIM による評価結果を表 1 に示す．表 1 より，(a) の台形歪みを加えた表画像と比較して，(b) の補正後の表画像では MS-SSIM の値が 23.76 ポイント向上した．この結果から，提案手法により表画像の幾何学的な歪みが緩和され，視覚的品質が改善されていることが確認できる．

次に，表構造解析精度への影響について述べる．台形歪みの補正についての実験結果を表 2 に示す．表 2 より，(b) の台形歪みを加えた表画像を入力した場合と比較して，(c) の台形歪みを補正した表画像を入力した場合，TEDS は 8.72 ポイント，S-TEDS は 9.89 ポイント改善しており，(a) の歪みのない表画像の TEDS および S-TEDS には及ばないものの，台形歪みによる精度低下を一定程度抑制できており，提案した補正手法が表構造解析精度の改善に有効であることが確認できた．

5.3.2 表構造解析結果の修正に関する実験の結果

LLM を用いた表構造解析結果の修正に関する実験結果を表 3 に示す．

まず，総合評価指標である TEDS に着目すると，基準となる修正なしの (a) が 0.9596 であるのに対し，RAG を使用せず LLM のみで修正を行った (b) および用語集のみを付与した (c) は，共に 0.9598 と微増した．さらに，外部知識として表構造情報を与えた (d) において，精度は 0.9599 となり，僅差ながら全条件の中で最高値を示した．これに対し，用語集と表構造の双方を組み合わせた (e) は 0.9594 に留まり，基準の (a) を下回る結果となった．

次に，表の構造的な正しさを評価する S-TEDS に着目する．(a) の 0.9699 に対し，(b) および (c) が 0.9708 と最も高い値を示し，次いで (d) が 0.9705 となった．(e) を除くすべての修正手法において，S-TEDS は基準を上回っており，LLM を用いた事後修正が構造的な誤りの訂正に寄与していることが確認できる．

S-TEDS では (b) や (c) が (d) を僅かに上回るものの，総合指標である TEDS では (d) が最高値を示した．これは，(d) が構造とテキスト内容の整合性を最も高い水準で両立できたためと考えられる．

5.4 考　　察

5.4.1 台形歪みを加えた表画像の補正に関する分析

台形歪みを加えた表画像の補正に関して，補正が失敗した事例を図 5 に示す．

図 5（c）に示すように，表の下部で検出された横方向の線分が，画像中に存在する物理的な罫線の長さを超えて画像の右端まで過剰に延長されている．この誤った線分検出により，表領域下端右側の境界座標が本来の位置よりも外側に推定された．その結果，消失点の推定および射影変換行列の算出に誤差が生じ，図 5（d）に示すような不適切な補正結果が得られた．

この過剰な線分延長の要因を分析する．対象画像の最下行に

表 1: 表画像の品質評価結果

手法	評価指標
	MS-SSIM
(a) 台形歪みを加えた表画像	0.3272
(b) 台形歪みを補正した表画像	**0.5648**

表 2: 表画像の表構造解析精度

手法	評価指標	
	TEDS	S-TEDS
(a) 元の表画像	0.9596	0.9699
(b) 台形歪みを加えた表画像	0.6333	0.7716
(c) 台形歪みを補正した表画像	**0.7205**	**0.8705**

表 3: LLM を用いて修正した表構造解析精度

手法	RAG の有無	評価指標	
		TEDS	S-TEDS
(a) 修正なし	RAG なし	0.9596	0.9699
(b) システムメッセージのみ		0.9598	**0.9708**
(c) 用語集のみ	RAG あり	0.9598	**0.9708**
(d) 表構造のみ		**0.9599**	0.9705
(e) 用語集+表構造		0.9594	0.9698

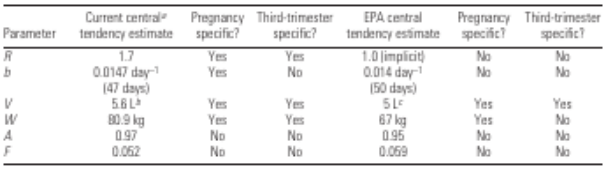
おいて，物理的な罫線は図 5 (b)に示すように表の右端付近で終端しているが，その右側には空白領域が存在する．この空白領域には，画像圧縮や輝度勾配などに起因する微細なノイズ成分が含まれており，これらがエッジ特徴として検出されていた．提案手法は，近接かつ近似した角度を持つ線分群を単一のクラスタに統合し，その両端点を結ぶ一本の線分として復元を行う．本事例では，座標系の歪みにより右端のノイズと物理的な罫線が幾何学的許容誤差内で同一線上に配置されたため，両者が誤統合された．その結果，本来の終端が無視され，画像全幅を貫通する線分が生成されたことが補正失敗の要因である．

5.4.2 LLM による表構造解析結果の修正に関する分析

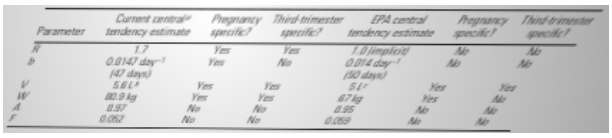
LLM を用いた修正処理が，表構造解析の評価指標である TEDS，および構造の正確性を測る S-TEDS に与える影響について，スコアが向上した事例および低下した事例を分析する．

まず，LLM を用いた修正処理により TEDS および S-TEDS が向上した事例を図 6 に示す．図 6 の事例では，修正後に TEDS スコアが 0.9154 から 0.9699 へと改善した．図 6(a) の表には，学歴区分を示す行として「<Technical」という文字列が含まれている．しかし，図 6(b) の TableMASTER による出力では，セル内の記号「<」が HTML タグの開始文字として誤って解釈され，構文エラーによって当該セル自体が構造から欠落する結果となっていた．これに対し提案手法では，LLM がセル内の文字列を文脈として再解釈した．その結果，図 6(c) の赤枠に示すように，「Technical」が教育水準を表す単一の項目であることを認識し，記号を除去した適切な文字列として再構成した．この修正により，セル内の文字列の一致度が高まったことで最終的な TEDS が向上した．また，構文エラーにより消失していたセルが表の木構造上に正しく復元されたため，構造の整合

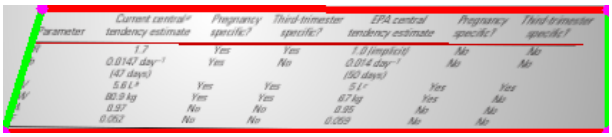
(a) 元の表画像



(b) 台形歪みを加えた表画像



(c) (b) の表画像に検出した表領域を重ねた表画像



(d) 台形歪みを補正した表画像

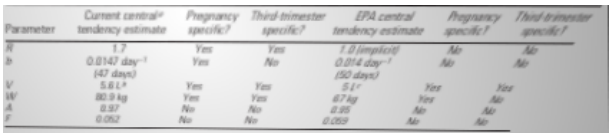


図 5: 台形歪みを加えた表画像の補正の失敗例（表画像の出典：ICDAR2021-SLP テストデータ）

性を示す S-TEDS の値も改善する結果となった．

一方，LLM を用いた修正により TEDS の値が低下した事例を図 7 に示す．図 7 の事例では，修正適用後に TEDS の値が 0.9891 から 0.8948 へと低下した．図 7(a) の元画像および図 7(b) の修正前の出力には，統計的な欠損値を表す「-」や，「$8.50E^{-01}$」のような指数表記が含まれている．提案手法において LLM は，セルの内容を一般的な数値形式へと適合させようとする過剰な正規化を行った．その結果，図 7(c) の青枠に示すように，LLM は欠損値を表す「-」を数学的な「0」へと置換し，さらに緑枠に示すように，指数表記の上付き文字タグ（<sup>）を削除した文字列へと改変した．本事例において，上付き文字タグの削除は表の行・列構成といった基本的な格子構造には影響しないため，S-TEDS の値は維持された．しかし，TEDS は正解データとの文字レベルでの厳密な一致度を評価する指標であるため，このような LLM による独断的な事実に基づかない出力や書式タグの欠落は，正解への忠実度の欠如とみなされ，TEDS スコアの大幅な低下を招いた．

以上の分析より，表構造解析における LLM を用いた修正処理に関して以下の知見が得られる．第一に，LLM は意味的文脈に基づく推論によって視覚的に曖昧な箇所のノイズを修復し，不適切なタグ解釈によるセルの欠落を防いで S-TEDS を向上させるなど，構造的な整合性を自律的に回復させる能力を有する．これは，局所的な画素情報に依存する従来の画像認識モデルの限界を，LLM の知識が補完する上で有効であることを示している．第二に，LLM の強力な推論能力は，専門的な記法

(a) 元の表画像

Social variable	Nonpregnant women	Pregnant women
Age (years)	29.250 ± 2.314	28.333 ± 1.971
Ethnicity	Han	Han
Occupation		
Housewife	0	3
Employee	20	27
Level of education		
<Technical	6	10
Bachelor	9	13
Master	5	7
Economic status	Regular	Regular

(b) 修正前の HTML コードから作成した表

Social variable	Nonpregnant women	Pregnant women
Age (years)	29.250 ± 2.314	28.333 ± 1.971
Ethnicity	Han 0.001	Han
Occupation		
Housewife	0	3
Employee	20	27
Level of education		
	6	10
Bachelor	9	13
Master	5	7
Economic status	Regular	Regular

(c) 修正後の HTML コードから作成した表

Social variable	Nonpregnant women	Pregnant women
Age (years)	29.250 ± 2.314	28.333 ± 1.971
Ethnicity	Han 0.001	Han
Occupation		
Housewife	0	3
Employee	20	27
Level of education		
Technical	6	10
Bachelor	9	13
Master	5	7
Economic status	Regular	Regular

図 6: LLM による表画像の修正で精度が向上した例（表画像の出典：ICDAR2021-SLP テストデータ）

を強制的に適合させようとする過剰な正規化のリスクを孕んでいる．今回の事例のように，S-TEDS が示す構造的な正しさは維持しつつも，文字情報の書き換えや書式タグの消去によって TEDS を低下させてしまうケースが確認された．したがって，実用化にあたっては，画像の忠実性を維持しつつ内容の改ざんを抑止するための安全策や制約条件の設計が不可欠であることを示唆している．

(a) 元の表画像（一部抜粋）

6	-	$8.50E^{-01}$
13	-	$6.80E^{-01}$
11	-	$5.90E^{-01}$
5	-	$5.40E^{-01}$

(b) 修正前の HTML コードから作成した表（一部抜粋）

6	-	$8.50E^{-01}$
13	-	$6.80E^{-01}$
11	-	$5.90E^{-01}$
5	-	$5.40E^{-01}$

(c) 修正後の HTML コードから作成した表（一部抜粋）

6	0	8.50E−01
13	0	6.80E−01
11	0	5.90E−01
5	0	5.40E−01

図 7: LLM による表画像の修正で精度が低下した例（表画像の出典：ICDAR2021-SLP テストデータ）

6 おわりに

本稿では，表画像に対する台形歪みの補正と大規模言語モデルを用いた表構造解析結果の修正を提案した．

台形歪みの補正では，人工的に台形歪みを加えた表画像データセットを構築した．その後，OpenCV を用いて生成した台形歪みを加えた表画像を補正することで，表構造解析モデルへの入力画像の品質および表構造解析精度の向上を実現した．LLM を用いた表構造解析結果の修正では，表構造解析モデルの出力として得られた表の HTML コードを専門用語集および表構造解析事例と併せて LLM に与えることで，セル中のテキストの誤りやタグの欠落や不整合などを修正した．

評価実験の結果，台形歪みの補正を適用することで，画像品質を示す MS-SSIM は 23.76 ポイント改善し 0.5648 へ，表構造解析精度を示す TEDS は 8.72 ポイント改善し 0.7205 となった．また，LLM を用いた解析結果の修正においては，外部知識として表構造情報を活用した際に TEDS が 0.03 ポイント向上し 0.9599 へ，S-TEDS が 0.06 ポイント向上し 0.9705 となった．

今後の課題としては，歪んだ表画像の補正において，台形歪みだけでなく，紙面の反りや折り目に起因する湾曲歪みや折れ歪みなどの非線形な歪みに対応することが挙げられる．また，表構造解析結果の修正に関しては，本稿では LLM として Llama-3.3-70B を用いたが，他の LLM を用いた場合の検証や，他の表構造解析モデルに対しても提案手法が有効であるかの検証などが挙げられる．さらに，台形歪みの補正時の画質劣化に起因する文字認識精度の低下を，LLM を用いた修正によって回復できるか検証することによる両手法の相乗効果の確認が挙げられる．

謝　辞

本研究の一部は，科学研究費補助金基盤研究 (B)(課題番号 23K25158) および 2025 年度国立情報学研究所公募型共同研究 (252FC-23662) の援助による．

文　献

[1] 納田 朋享, 金沢 輝一, 上野 史, 太田 学, “表画像を入力とする表構造解析手法の分析と改良,” 第 17 回データ工学と情報マネジメントに関するフォーラム (DEIM 2025), 4G-03, 2025.

[2] 田上 歩夢, 金澤 輝一, 上野 史, 太田 学, “表構造情報を利用した棒グラフの自動生成の一手法,” 第 16 回データ工学と情報マネジメントに関するフォーラム (DEIM 2024), T4-A-3-02, 2024.

[3] Hiroyuki Shindo, Yuji Matsumoto, Masashi Ishii, Hiroyuki Oka, Atsushi Yoshizawa, “Machine extraction of polymer data from tables using XML versions of scientific articles,” *Science and Technology of Advanced Materials: Methods*, Volume 1, pp. 11–23, 2021.

[4] Jiaquan Ye, Xianbiao Qi, Yelin He, Yihao Chen, Dengyi Gu, Peng Gao, Rong Xiao, “PingAn-VCGroup’s Solution for ICDAR 2021 Competition on Scientific Literature Parsing Task B: Table Recognition to HTML,” arXiv preprint arXiv:2105.01848, 2021.

[5] Nam Tuan Ly, Atsuhiro Takasu, “An End-to-End Multi-Task Learning Model for Image-based Table Recognition,” *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP*, pp. 626–634, 2023.

[6] Rujiao Long, Wen Wang, Nan Xue, Feiyu Gao, Zhibo Yang, Yongpan Wang, Gui-Song Xia, “Parsing Table Structures in the Wild,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Volume 1, pp. 924–932, 2021.

[7] Gavin Greif, Niclas Griesshaber, Robin Greif, “Multimodal LLMs for OCR, OCR Post-Correction, and Named Entity Recognition in Historical Documents,” arXiv preprint arXiv:2504.00414, 2025.

[8] Ahmed Masry, Do Xuan Long, Jia Qing Tan, Shafiq Joty, Enamul Hoque, “ChartQA: A Benchmark for Question Answering about Charts with Visual and Logical Reasoning,” *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 2263–2277, 2022.

[9] Minesh Mathew, Dimosthenis Karatzas, C.V. Jawahar, “DocVQA: A Dataset for VQA on Document Images,” 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 2199–2208, 2021.

[10] Antonio Jimeno Yepes, Peter Zhong, Douglas Burdick, “ICDAR 2021 Competition on Scientific Literature Parsing,” *Proceedings of 16th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 605–617, 2021.

[11] Ning Lu, Wenwen Yu, Xianbiao Qi, Yihao Chen, Ping Gong, Rong Xiao, Xiang Bai, “MASTER: Multi-Aspect Non-local Network for Scene Text Recognition,” *Pattern Recognition*, Volume 117, Article number 107980, 2021.

[12] Wenhai Wang, Enze Xie, Xiang Li, Wenbo Hou, Tong Lu, Gang Yu, Shuai Shao, “Shape robust text detection with progressive scale expansion network,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9336–9345, 2019.

[13] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko, “End-to-End Object Detection with Transformers,” *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 213–229, 2020.

[14] Alex Smock, Rohith Anil, Mark Hasegawa-Johnson, Matthew A. Gardner, “PubTables-1M: Towards Comprehensive Table Extraction from Unstructured Documents,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4634–4642, 2022.

[15] Hmrishav Bandyopadhyay, Tanmoy Dasgupta, Nibaran Das, Mita Nasipuri, “A Gated and Bifurcated Stacked U-Net Module for Document Image Dewarping,” 2020 25th International Conference on Pattern Recognition (ICPR), pp. 10548–10554, 2021.

[16] Ke Ma, Zhixin Shu, Xue Bai, Jue Wang, Dimitris Samaras, “DocUNet: Document Image Unwarping via a Stacked U-Net,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4709, 2018.

[17] Ziyi Zhu, Zhi Tang, Liangcai Gao, “Table image dewarping with key element segmentation,” *International Journal on Document Analysis and Recognition (IJDAR)*, Volume 27, pp. 349–362, 2024.

[18] Yi Ren, Chenglong Yu, Weibin Li, Wei Li, Zixuan Zhu, Tianyi Zhang, Chenhao Qin, Wenbo Ji, Jianjun Zhang, “TableGPT: a novel table understanding method based on table recognition and large language model collaborative enhancement,” *Applied Intelligence*, Volume 55, Article number 311, 2025.

[19] Tianshu Zhang, Xiang Yue, Yifei Li, and Huan Sun, “TableLlama: Towards Open Large Generalist Models for Tables,” *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pp. 4361–4378, 2024.

[20] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wentau Yih, Tim Rocktäschel, Sebastian Riedel, Douwe Kiela, “Retrieval-augmented generation for knowledge-intensive NLP tasks,” *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 9459–9474, 2020.

[21] Xu Zhong, Elaheh ShafieiBavani, Antonio Jimeno Yepes, “Image-based table recognition: data, model, and evaluation,” *Computer Vision – ECCV 2020*, pp. 564–580, 2020.

[22] Yongshuai Huang, Ning Lu, Dapeng Chen, Yibo Li, Zecheng Xie, Shenggao Zhu, Liangcai Gao, Wei Peng, “Improving Table Structure Recognition with Visual-Alignment Sequential Coordinate Modeling,” 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11134–11143, 2023.

[23] Zhou Wang, Eero P. Simoncelli, Alan C. Bovik, “Multiscale structural similarity for image quality assessment,” *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*, Volume 2, pp. 1398–1402, 2003.