CH-003

固定カメラ画像に対する降雨強度推定に向けた注視改善を伴う回帰モデルの構築

Development of an Attention-Enhanced Regression Model for Rainfall Intensity Estimation from Fixed-Camera Images

矢野 耕太郎 † 遠藤 聡志 ‡ 武井 弘樹 §
Kotaro Yano† Satoshi Endo‡ Hiroki Takei§

1 はじめに

沖縄県は、地理的・気象的な特性により、局所的かつ突発的な豪雨が多発する地域である。こうした豪雨は都市部の内水氾濫など深刻なインパクトをもたらしており、そのリスク低減には「いつ」「どこで」「どれだけ」雨が降っているかを精緻に把握することが不可欠である。しかし、従来の数値予報モデルや観測網だけでは、きめ細やかな現場の実況を十分に捉えきれないという課題があった。

このギャップを埋めるため、ウェザーニューズはユーザー参加型のウェザーリポートを展開している。末光ら [1] はこの仕組みを応用し、画像から天気を自動判別するシステムを提案したが、定性的な分類にとどまり、防災上より重要な定量的な降雨強度推定には至っていなかった。

そこで本研究では、単一の固定カメラ画像から降雨強度（mm 単位）を推定する深層学習モデルの構築を目指す。先行研究では CNN による回帰モデルが提案されているが [2]、特に強雨領域で推定精度が著しく低下する課題があった。

この課題に対し、年齢推定分野で有効性が示された [3]coarse-to-fine 戦略（大まかなカテゴリ分類から詳細な数値推定への段階的アプローチ）を採用し、強雨領域の精度向上を図る。

さらに初期検討で、一般的な CNN では注視領域が雲の状態と無関係な箇所に偏る場合があることが分かった。そこで本研究では、画像全体の文脈を捉える Vision Transformer (ViT) を特徴抽出器とし、モデルの注視点を本質的な領域に誘導することで、解釈性と精度の両立を図る。

2 提案手法

本研究では、単一の固定カメラ画像から現在の降雨強度を高精度に推定するため、分類と回帰を段階的に組み合わせたカスケード型のモデルを提案する。図 1 は、本研究で提案する手法と、比較対象となる先行研究のプロセスの違いを模式的に示したものである。 本手法は、特に先行研究で課題とされていた強い雨の推定精度を克服することを目的とする。

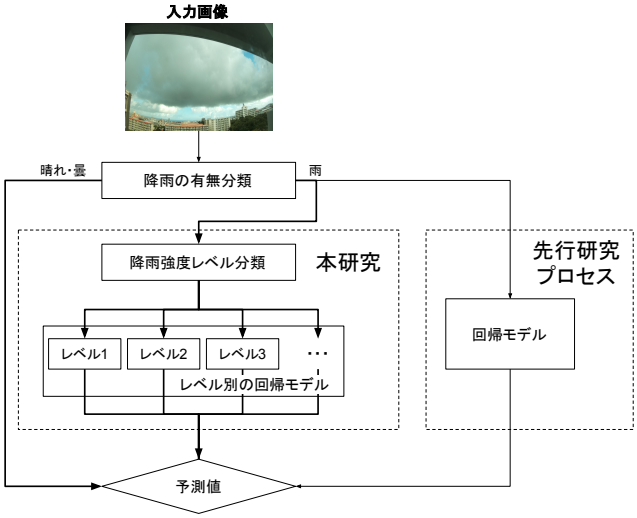


図 1: 本研究の構成モデルのイメージ図

図 1 に示すように、まず入力画像に対し「晴れ/雨」の 2 値分類を行い、降雨の有無を判断する。このステップは両プロセスで共通である。先行研究では、ここで「雨」と分類された画像に対し、単一の回帰モデルを直接適用して降水量を推定していた。

これに対し本研究の提案手法は、Chen ら [3] が提案した coarse-to-fine 戦略に基づき、回帰の前に降雨強度レベル分類というステップを新たに導入する。このステップでは、降雨強度という連続的な値を、あらかじめ定義した「レベル 1」「レベル 2」といった複数の離散的なレベルに一旦分類する。

さらに、分類された各強度レベルに対して、それぞれ専用の回帰モデルを構築する。これにより、「強い雨」のデータのみを学習した回帰モデルや、「弱い雨」に特化した回帰モデルなど、各モデルがより単純化されたタスクに専念することが可能となり、全体の予測精度向上を目指す。本研究では、後段の回帰精度に最も寄与する最適なレベル分類数を探索するため、複数のパターンで精度を比較評価する。

なお、これら一連のプロセスにおける特徴抽出器には、はじめにで述べた注視点の課題を解決するため、ViT を採用する。CNN が局所的な特徴を捉えるのに対し、ViT は画像全体の大域的な文脈を捉えることができる。この特性により、モデルが降雨強度と本質的に関連の深い領域に注視することを期待する。

† 琉球大学大学院理工学研究科知能情報プログラム,Graduate School of Engineering and Science, University of the Ryukyus
‡ 琉球大学工学部工学科知能情報コース, Computer Science and Intelligent Systems, University of the Ryukyus
§ 株式会社ウェザーニューズ, Weathernews Inc.

Copyright © 2025 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

3 データセット

本研究では、単一の雲画像から現在の降雨強度を高精度に推定するモデルの構築を目指している。このモデルの学習および評価のためには、様々な気象条件下で撮影された雲画像と、それぞれの画像撮影時刻における正確な降雨強度の実測値が必要となる。そこで本研究では、定点カメラ等で撮影された雲画像に対し、気象レーダーによって観測された高解像度の雨量データを正解ラベルとして付与したデータセットを新たに構築した。図 2 にデータセットの構築の手順について示す。

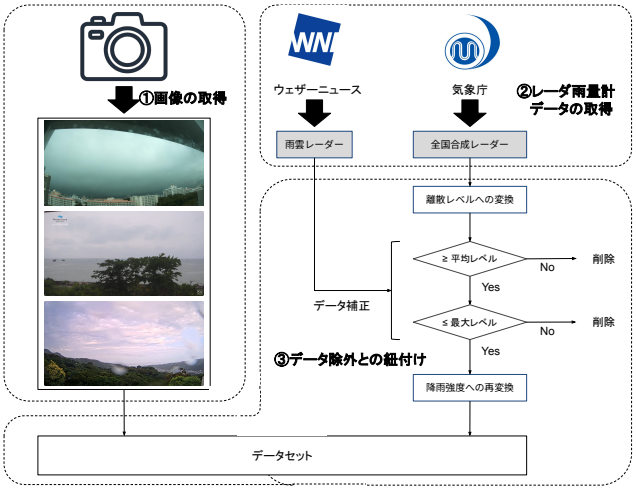


図 2: データセット構築の概要

3.1 画像の取得

図 2 の「①画像の取得」について, 画像は沖縄県西原町に設置した固定カメラと、許諾を得たのライブ配信の 2 つの方法で取得した。固定カメラでは北西方向の空を 7 時から 18 時まで 1 分毎に撮影した。YouTube 配信は宮古島市伊良部島 [?] と国頭郡本部町 [?] の映像を用い、画面の約半分が空を映すものを対象として 5 分毎にスクリーンショットを取得した。

3.2 レーダ雨量計データの取得

図 2 の「②レーダ雨量計データの取得」について, レーダ雨量計データは、京都大学の「全国合成レーダー GPV」と、ウェザーニューズ社の「雨雲レーダー」の 2 種類を利用した。合成レーダーからは、10 分毎（2024 年 3 月以降は 5 分毎）に定量的な降雨強度を取得した。一方、雨雲レーダーからは、5 分毎に色で表現された降雨強度レベルを取得した。具体的には、カメラ撮影地点周辺のレーダー画像をスクリーンショットで保存し、画像内のピクセル色を解析してレベルに変換後、エリア内の最大レベルと最頻レベルを記録した。

3.3 データ除外と紐付け

図 2 の「③データの除外と紐付け」について説明する. まず、2 種類のレーダーデータを時刻基準で紐付け、信頼性の低いデータを除外した。より高精度な雨雲レーダーの降雨強度レベルを正とし、合成レーダーの値をレベル変換した際に、その値が雨雲レーダーのレベル範囲（最頻値〜最大値）に含まれないデータを不整合として除外した。次に、このフィルタリング後のレーダーデータと画像を紐付けた。両者には最大 2 分 30 秒の時刻差を許容し、固定カメラは同一時刻、YouTube 画像は最も時刻が近いものを対応させた。以上の手順により、雲画像に合成レーダーの降雨強度値をラベルとして付与したデータセットを構築した。

4 実験概要

本章では、提案モデルの有効性を検証するために実施した実験のセットアップについて述べる。まず、性能評価に用いる指標を定義し、次に降雨強度レベル分類の定義を説明する。さらに、比較対象モデルのアーキテクチャと学習パラメータを示し、最後にモデルの判断根拠を分析する可視化手法について解説する。

4.1 評価指標

本研究ではモデルの汎化性能を安定して評価するため、5 分割交差検証を採用し、評価指標には主に平均絶対誤差（MAE）と最大誤差（MaxE）を用いる。降雨強度データは値の範囲が広く、裾の重い分布を持つため、二乗誤差（MSE）は少数の極端な誤差の影響を受けやすい。そこで、外れ値に頑健で直感的に解釈しやすい MAE を主要な指標とする。一方で、MAE だけではモデルの致命的な予測誤差を評価できないため、防災応用を考慮し、最悪ケースも把握する補助指標として MaxE を併用し、モデルの安定性と信頼性を多角的に評価する。

4.2 降雨強度レベルの定義

提案モデルの降雨強度レベル分類における分類数と降雨強度の関係を表 1 に示す。分類数は 3, 4, 5, 6, 7, 13 で、約 0.1〜10mm/h を細かく分類し、10mm/h 以上はデータ数が少ないため 1 クラスに統合している。これらの分類は、気象庁やウェザーニューズなど各機関の基準を参考に、データ数も考慮して再定義した。また、レベル 1〜13 は雨の強さを区別し、晴れ・曇りはレベル 0 として扱う。

4.3 アーキテクチャと学習パラメータ

本研究で提案するモデルの特徴抽出器には、ImageNet で事前学習済みの Vision Transformer(ViT-Base/16) モデルを採用する。転移学習の効率化のため、ViT の前半の層を凍結し、その上に最終的な分類や回帰を担う MLP 層を新たに追加した構造を用いる。

学習においては、最適なハイパーパラメータを探索した。探索対象は、バッチサイズ、データ拡張のサイズ、および正則化のための weight decay の 3 つの組み合わせである。学習の最適化には AdamW オプティマイザを用い、分類タスクの損失関数には Focal Loss を採用した。また、学習の進捗に応じて学習率を動的に調整する学習率スケジューラを併用し、検証用データの損失が 15 エポック連続で改善しなかった場合に学習を打ち切る Early Stopping も導入している。

Copyright © 2025 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

表 1: 提案モデルにおける多値分類の降雨強度レベルと降雨強度の関係

降雨強度レベル	1	2	3	4	5	6	7	8	9	10	11	12	13
3 値分類の降雨強度幅 (mm/h)	0.1∼5	5∼10	10∼										
4 値分類の降雨強度幅 (mm/h)	0.1∼1	1∼5	5∼10	10∼									
5 値分類の降雨強度幅 (mm/h)	0.1∼1	1∼2	2∼4	4∼8	8∼								
6 値分類の降雨強度幅 (mm/h)	0.1∼1	1∼2	2∼4	4∼6	6∼10	10∼							
7 値分類の降雨強度幅 (mm/h)	0.1∼0.5	0.5∼1	1∼2	2∼3	3∼5	5∼10	10∼						
13 値分類の降雨強度幅 (mm/h)	0.1∼0.25	0.25∼0.5	0.5∼0.75	0.75∼1	1∼1.5	1.5∼2	2∼2.5	2.5∼3	3∼4	4∼5	5∼6	6∼10	10∼

4.4 可視化手法

モデルの判断根拠を分析するため、Grad-CAM を用いる。これは予測に対する勾配情報を利用し、判断に寄与した画像領域を可視化する手法である。本研究で採用する ViT への Grad-CAM の適用は、Chefer ら [6] によりその有効性が示されている。

本研究ではこの可視化を通じ、モデルが雲の形状など物理的に意味のある特徴を捉えているか、また予測誤差の原因はどこにあるかを分析する。これにより、モデルの信頼性を質的に評価し、性能改善の知見を得ることを目指す。

5 結果と考察

5.1 EfficientNet と ViT の精度評価

表 2 に、Coarse-to-fine 戦略を採用した本提案モデル、および比較対象である直接回帰アプローチの MAE と MaxE を示す。

表 2: EfficientNet と ViT の性能比較（MAE ／ MaxE）

モデル	EfficientNet		ViT	
	MAE	MaxE	MAE	MaxE
3 値 + 回帰	1.68	44.01	1.85	65.40
4 値 + 回帰	1.60	58.20	1.65	65.65
5 値 + 回帰	1.60	56.12	1.73	102.61
6 値 + 回帰	1.58	59.62	1.75	87.85
7 値 + 回帰	1.57	50.88	1.68	68.56
13 値 + 回帰	1.63	42.10	1.95	82.99
先行研究プロセス	1.88	50.63	1.47	57.67

表より最も注目する点は、モデルのアーキテクチャとアプローチの組み合わせによって優位性が逆転することである。提案モデルにおいて、ViT は全てのクラス数で EfficientNet よりも高い MAE を示し、特に MaxE が著しく悪化した。EfficientNet で最も MAE が良かった 7 値 + 回帰では、MAE=1.57 から MAE=1.68 へと約 0.1mm の精度の悪化が見られ、MaxE に関しても 50.88mm から 68.56mm と 20mm 近くの悪化が見られる。対照的に、先行研究プロセスでは ViT の MAE が EfficientNet を大きく下回り、優れた性能を示した。この結果は、なぜ ViT の性能が採用するアプローチによって逆転するのかという問いを提起する。

提案モデルにおける性能逆転の要因を解明するため、Coarse-to-fine 戦略の前段を担う分類ステージの挙動を、エラーの質という観点から分析する。図??に示す両モデルの混同行列は、そのエラーパターンの違いを明確に示している。EfficientNet の誤分類は対角線近傍に集中しており、誤差が近しいレベルに収まる傾向がある。一方、ViT は特定のレベルへ誤分類が偏る傾向が見られ、真のレベルから大きくかけ離れた間違いを犯す場合がある。このエラーは、Coarse-to-fine という枠組みにおいて、後段の専門化された回帰モデルとの深刻なミスマッチを引き起こす。表 3 が示す通り、分類が正解した場合の MAE は両モデルで大差ないが、誤分類した場合には ViT の MAE が著しく増大し、この仮説を裏付けている。

(a) EfficientNet の混同行列　(b) ViT の混同行列

図 3: 7 値分類における EfficientNet と ViT の混同行列

5.2 可視化によるモデルの挙動分析

5.2.1 EfficientNet と ViT の比較

本研究で EfficientNet から ViT へモデルを変更した主要な目的は、モデルが画像中のどこに注目して判断を下しているか、その可視化結果を改善することにある。図 4 に両モデルの注目領域の可視化結果を示す。図 4a の EfficientNet の可視化結果では、画像の左上に存在するマスク領域に対して、モデルが強く注目していることが確認できる。これは、本来降雨量予測とは無関係であるべき画像編集の痕跡に、モデルの判断が影響されている可能性を示唆する。一方で、図 4b の ViT の可視化結果では、マスク領域への注目は大幅に抑制され、むしろその周辺領域や空模様、海面の様子など、より広範な文脈に注意が分散している。この結果から、ViT を導入することで、マスク領域や画像の四隅といった本質的でない情報への過度な注目を避け、より妥当な領域に基づいて判断を行えるよう、モデルの挙動が改善されたことがわかる。

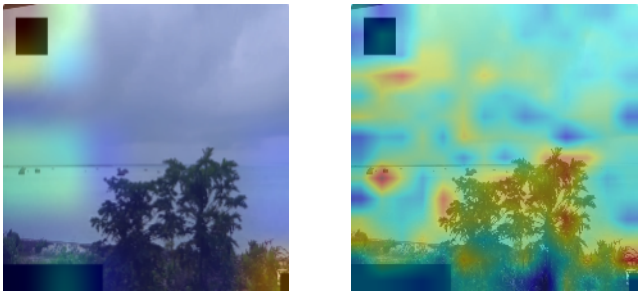
5.2.2 提案モデルと先行研究プロセスの比較

次に、同じ ViT アーキテクチャ内でも、Coarse-to-fine 戦略と直接回帰というアプローチの違いが、モデルの注目領域にどのような影響を与えるかを具体的な事例で分析する。

Copyright © 2025 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

表 3: 正分類・誤分類時における性能比較

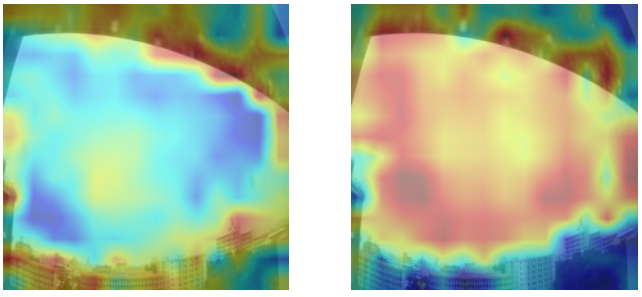
降雨量範囲 (mm)	EfficientNet (正分類)	ViT (正分類)	EfficientNet (誤分類)	ViT (誤分類)
0.1∼0.5	0.07 / 0.17	0.08 / 0.22	1.00 / 6.48	1.38 / 6.95
0.5∼1	0.14 / 0.42	0.13 / 0.38	0.97 / 8.84	1.09 / 17.65
1∼2	0.26 / 0.83	0.25 / 0.56	1.47 / 14.70	1.60 / 13.78
2∼3	0.28 / 0.69	0.17 / 0.67	2.40 / 19.50	2.43 / 17.11
3∼5	0.58 / 1.51	0.61 / 1.63	3.16 / 39.53	3.86 / 25.26
5∼10	1.31 / 3.87	1.20 / 3.20	5.70 / 30.44	6.28 / 18.33
10∼	13.08 / 50.88	14.23 / 68.56	14.26 / 36.18	13.35 / 50.67
全体 (Overall)	**2.44 / 50.88**	**2.25 / 68.56**	**2.66 / 39.53**	**3.08 / 50.67**



(a) EfficientNet の可視化　(b) ViT の可視化

図 4: EfficientNet と ViT の可視化結果

図 5 は、同じ入力画像（真値: 2.75）に対する先行研究プロセスと提案モデル（7 値分類 + 回帰）の可視化結果を示している。図 5a の直接回帰モデルでは、空の中心部に加え、下部の建物群にも注意が分散していることがわかる。このモデルは比較的広範な情報を基に判断し、予測値 6.60mm（誤差 3.85mm）を出力した。対照的に図 5b の提案モデルでは、注目領域が空の中心部に極めて強く集中しており、建物などの他の要素はほとんど無視されている。この「空」という単一の情報源への過度な集中が、結果として 12.49mm という極端な過大予測を招き、誤差を 9.74mm へと倍以上に増大させる一因となったと考えられる。この事例は、直接回帰モデルがより大局的な文脈から判断を下すのに対し、提案モデルは特定の情報源に集中するあまり極端な予測を導く可能性があるなど、両アプローチでモデルの判断プロセスに質的な違いがあることを示唆している。



(a) 先行研究プロセス　(b) 7 値分類 + 回帰

図 5: 先行研究プロセスと提案モデルの可視化結果

6　まとめと今後の展望

本研究では、Coarse-to-fine 戦略と ViT を組み合わせた降雨強度推定モデルを提案した。ViT は CNN より妥当な領域に注目したが、提案モデルはデータ不足に起因する分類エラー時の誤差が著しく増大した。この結果は、本戦略を ViT で有効に機能させるには、十分なデータに基づいた頑健な分類器が不可欠であることを示している。今後の展望として、学習データセットの物理的な拡充や、AugMix 等の高度なデータ拡張を適用し、特にデータが少ない強雨クラスの頑健性を高めることで、さらなる精度向上が期待される。

7　謝辞

本研究は JSPS 科研費 23K11234 の助成を受けたものである.

参考文献

[1] K. Suemitsu, et al,:”Selection of Dash Cam Images for Weather Forecasting Based on The Sky Occupancy,” 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS), Ise, Japan, 2022, pp. 1-8, doi: 10.1109/SCISISIS55246.2022.10002033.

[2] J. Byun, et al,:”Deep Learning-Based Rainfall Prediction Using Cloud Image Analysis,” in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-11, 2023, Art no. 4701411, doi: 10.1109/TGRS.2023.3263872.

[3] Chen, J. C., et al (2016, September). A cascaded convolutional neural network for age estimation of unconstrained faces. In 2016 IEEE 8th International conference on biometrics theory, applications and systems (BTAS) (pp. 1-8). IEEE.

[4] Youtube,KuROKO-宮古島映像-, https://www.youtube.com/live/4v5e4eKIT_E?si=jkyHDaMiYcg6_6p,2024/06

[5] Youtube,Motobu Terrace, https://www.youtube.com/live/Tlc3uegvdrQ?si=xMDiI4OiMQKkalO0,2024/06

[6] Chefer, H., et al,: (2021). Transformer interpretability beyond attention visualization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 782-791).

Copyright © 2025 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.