



Title	Disaster Recognition Through Image Captioning Features and Shifted Attention
Author(s)	Thanyawet, Narongthat
Citation	大阪大学, 2024, 博士論文
Version Type	VoR
URL	<a href="https://doi.org/10.18910/101462">https://doi.org/10.18910/101462</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## Abstract of Thesis

Name (THANYAWET Narongthat)	
Title	Disaster Recognition Through Image Captioning Features and Shifted Attention (キャプショニングとアテンションシフト機構に基づく災害画像認識)

### Abstract of Thesis

In the wake of escalating natural disasters, timely and precise recognition of affected areas has become imperative for efficient disaster management and mitigation. Traditional image processing techniques often fall short in identifying subtle nuances within disaster-stricken regions due to their propensity to highlight prominent features, thereby overlooking critical details. This thesis presents a method that utilized image captioning features along with adaptive attention mechanisms to improve the recognition of disaster-affected regions.

The author proposes a novel methodology that integrates image captioning with a custom-developed attention-shifting algorithm designed to dynamically refocus the model on less conspicuous yet essential elements within images. By leveraging the inherent strengths of Vision Encoder-Decoder (VED) models, along with innovative optimal masking strategies, we enable the system to discern and articulate the specifics of disaster impacts in diverse imaging conditions, from satellite to ground-level perspectives.

The research methodology includes the rigorous training and evaluation of the model using extensive datasets comprising side-view, aerial, and shipborne images of disaster scenes. The model's performance is assessed against standard metrics, demonstrating a significant leap in accuracy and contextual relevance of the generated captions.

The empirical results underscore the superiority of our approach over conventional image captioning models, exhibiting enhanced detection capabilities with accuracies exceeding 91% for landslide detection from side-view image captions and 87.5% for shipborne view detection. These figures not only reflect the technical prowess of the system but also its practical applicability in real-world disaster assessment scenarios.

This work carries profound implications for the field of disaster management. By augmenting the quality and reliability of disaster region identification, our framework facilitates more informed decision-making in allocating resources for relief efforts. Additionally, the adaptive nature of the model paves the way for its application across a spectrum of environmental monitoring and emergency response tasks, heralding a new era of AI-enabled disaster management tools. Future research avenues include scaling the model to encompass a broader range of disaster types and integrating real-time data for swift, actionable insights during crisis events.

## 論文審査の結果の要旨及び担当者

氏 名 (Narongthat THANYAWET)		
	(職)	氏 名
論文審査担当者	主 査	教授 浦西 友樹
	副 査	教授 土屋 達弘
	副 査	教授 河原 吉伸
	副 査	准教授 白井 詩沙香

## 論文審査の結果の要旨

本博士論文は、アテンションメカニズムを用いた機械学習モデルが、画像内における異常をどのように理解できるかを焦点としている。具体的には、機械学習が災害状況、土砂崩れや洪水といった異常な状況をどのように正確に識別できるかを取り扱う。

第3章ではまず、アテンションメカニズムを用いた機械学習モデルが、従来のConvolutional Neural Network (CNN) のような非アテンションメカニズムベースのモデルと比べ、性能面でどの程度優れているか調査している。CNNは対象物そのものに注目し、周辺のコンテキストを考慮しない。このため、たとえCNNを使って土砂崩れ画像を用い学習した場合でも、正常な土壤と識別できず、誤検出率が上昇する問題がある。これに対し、アテンションメカニズムを活用することで、画像内の主要な領域を検出すると同時に周辺のコンテキストも考慮できる。提案手法ではこのモデルを使用して画像のキャプションを生成し、異常検出を行う。例えば土砂崩れの場合、たとえキャプションとして直接「土砂崩れ」という単語を生成できなくても、「土」や「滑る」といったキーワードが生成され、結果的にその画像が異常な状況を撮影したものであると分類できる。提案手法を用いて「土」や「滑る」といったキーワードが正しく生成されるのは、提案するモデルがアテンションメカニズムによって画像内の関係性を検出し、画像の構造を理解する能力を有していることに起因する。例えば前述の場合については、土が地面ではなく山に存在し、結果として滑り落ちていることを理解できる。

しかしながら、アテンションメカニズムは画像内の関係性を理解する能力がある一方で、特定の状況下で重要な情報を見逃す可能性も存在する。例えば画像内の災害領域が小さい場合や他の大きな物体に遮蔽されている場合には、生成されるキャプションが災害領域について言及しないことがある。このような場合においては、アテンションメカニズムが目立つ対象物に注目しすぎることで、結果として重要な異常を見逃す可能性がある。

そこで、本論文の第4章では「Detective Network」というモデルを提案している。このモデルは、画像内で隠れている重要な領域を見つけ出し、異常検知のために小さな領域を特定する役割を果たす。具体的には、大きな物体を適切にマスクすることでアテンションをシフトさせ、より小さな領域に関連するキャプションを生成できる。結果的に、提案するモデルを用いることで画像内の小さな異常を検出可能となり、従来手法に比べて精度や再現率が大幅に向上した。

これらの結果は、災害現場における飛行物体からの空撮をはじめとして、数多くの場面において災害検出の精度を高めることができることを示しており、顕著な工学的価値を有していることは疑いない。よって、本論文は博士論文（情報科学）の学位論文として価値のあるものと認める。