（ 続紙 1 ）

京都大学	博士（情報学）	氏名	蔡　溯東（CAI SUDONG）
論文題目	Learning Discriminative Neural Representations for Visual Recognition （画像認識のための識別性の高いニューラル表現の学習）		

（論文内容の要旨）

This thesis focuses on the derivation of novel activation functions for learning effective neural representations mainly for visual recognition tasks. The thesis re-interprets neural activation as feature selection and, through the investigation of adaptive neural feature selection at different levels (i.e., feature descriptors and units), proposes novel activation functions for visual recognition.

The thesis motivates the re-interpretation by examining the task of RGB-image road-scene material segmentation (RMS). The same road scene image can simultaneously have different regions of object categories and material categories, which signifies the importance of effective feature selection tailored to the task. In contrast to the original interpretation of activation functions as biological activation potential in the primate's neural network, this thesis proposes to interpret them as part of Multi Criteria Decision Making (MCDM), that is, as selective feature re-calibrators that suppress and emphasize features according to their importance measured by feature-filter similarities. Based on this new interpretation, novel activation models are derived.

The thesis presents the derivation of the new activation functions, the construction of frameworks for evaluating their effectiveness, the results of experimental comparisons with related methods, and discussion on future directions of research. The thesis is structured into 6 chapters.

In Chapter 1, the thesis discusses the importance of effective feature selection in neural representation learning for visual recognition. Through qualitative discussion on the task of RGB RMS, the thesis argues for the need of effective encoding and fusion of multi-scale texture cues and image context. For this, a novel self-attention mechanism with linear complexity is derived from regular spatial self-attention. The chapter also introduces the MCDM interpretation and outlines the derivation of novel activation functions based on this interpretation. Finally, the contributions of this research are clarified.

In Chapter 2, background knowledge and techniques that form the basis for the proposed models are introduced. This includes basic self-attention and self-gating mechanisms for image inputs and fundamentals of MCDM methods including TOPSIS, Fuzzy Comprehensive Evaluation, and Grey Relational Analysis.

In Chapter 3, a novel self-attention-based framework that effectively fuses texture and context cues of road scene materials is introduced for RGB RMS. The framework, RMSNet, encodes multi-scale multi-level features with hierarchical Transformer layers and leverages "SAMixer" for efficient context-aware multi-scale feature fusion. SAMixer, is built on two new modules: Efficient Balanced Multi-head Self-Attention (E-B-MSA) which extends spatial multi-head self-attention to N-to-1 feature fusion with linear complexity, and Bottleneck Local Statistics Encoding-Decoding (BLSED) to incorporate local details while accelerating E-B-MSA computation. Extensive experimental evaluations and ablation studies are shown to validate the effectiveness and generalizability of the proposed model.

In Chapter 4, an MCDM interpretation of neural activation functions is introduced. The thesis argues that this interpretation makes clear that a network can have an inherent problem in the feature selection, referred to as Mismatch Feature Scoring (MFS) due to the fact that large feature and filter norms can significantly bias inner-product-based similarity evaluation.

Based on this argument, the chapter introduces the Instantaneous Importance Estimation Units (IIEU) based on adaptive norm-decoupled feature-filter similarities. Experimental results demonstrate the improvements of IIEUs over past activation models.

In Chapter 5, the thesis introduces "AdaShift," an adaptive shift factor in the activation function, which dynamically translates the activation inputs by jointly leveraging informative local and non-local cues of different ranges. It enables fine-grained adjustments of the nonlinearities for soft feature selection. Experimental results demonstrate the effectiveness of AdaShift over other activation models.

In Chapter 6, a summary of the objectives and proposed methods of this thesis is provided, along with a discussion on future directions and applications.

（続紙 ２ ）

（論文審査の結果の要旨）

本論文は画像認識などに代表される視覚情報処理のための深層学習モデルに用いる新たな活性化関数の導出を行っている。活性化関数を複数基準を用いた意思決定問題（Multi-Criteria Decision Making）の一部として解釈することを提唱し、効率が良く効果的な特徴選択および統合を行う活性化関数の導出を行っている。本論文ではこれら新たな活性化関数の解釈と導出をRGB画像からの道路状況における素材認識問題を動機として議論し、特に、自己注意機構並びに自己ゲート機構を拡張することにより、画像全体の情報を組み込んだ特徴量選択および統合を行う活性化関数を提案している。本論文では、これらの新たな動的活性化関数の定量的評価および従来手法との比較を大規模画像認識データセット等を用いて行っている。本論文で得られた成果は以下の通りである。

（１）　RMSNetと呼ばれる、空間的自己注意機構を拡張することによりマルチスケール特徴量を効率的に学習および活用した、RGB画像を用いた道路状況における素材認識のための新たな深層学習モデル。

（２）　活性化関数を複数基準を用いた意思決定問題（Multi-Criteria Decision Making）の一部として解釈することにより、特徴量選択を担う関数として定式化し、新たな活性化関数の系統立てた導出を実現。

（３）　Instantaneous Importance Estimation Units（IIEU)と呼ぶ、適応的に大きさに寄らずに特徴量同士の似通い度から選択を行う新たな活性化関数とその有用性の画像認識における実証。

（４）　AdaShiftと名付けた、活性化関数のバイアスを局所および非局所特徴量に依存して動的に変化させる新たな手法の導出とその実験的有効性の実証。

以上、本論文は、主に視覚情報処理のための新たな活性化関数の導出を行い、それらの有効性を大規模画像認識データセット等で実証したもので、学術上、実際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、令和６年２月２１日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。また、本論文のインターネットでの全文公表についても支障がないことを確認した。