Thesis for Bachelor's Degree

Inductive bias provision for color attention learning

Park, Jiwon (박 지원 朴 支元)

School of Electrical Engineering and Computer Science

Gwangju Institute of Science and Technology

2023

Inductive bias provision for color attention learning

귀납 편향 제공을 위한 색채 어텐션 학습

Inductive bias provision for color attention learning

Advisor : Professor Kim, Sundong

by

Park, Jiwon

School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology

A thesis submitted to the faculty of Gwangju Institute of Science and Technology in partial fulfillment of the requirements for the degree of Bachelor of Science in Electrical Engineering and Computer Science

Gwangju, Republic of Korea

2023. 12. 18.

Approved by

Professor Kim, Sundong

Committee Chair

Inductive bias provision for color attention learning

Park, Jiwon

Accepted in partial fulfillment of requirements
for the degree of Bachelor of Science

December. 18. 2023.

Committee Chair _____
Prof. Sundong Kim

Committee Member _____
Prof. Chang Wook Ahn

BS/EC
20205088

Park, Jiwon (박 지원). Inductive bias provision for color attention learning (귀납 편향 제공을 위한 색채 어텐션 학습). School of Electrical Engineering and Computer Science. 2023. 9p. Advisor Prof. Kim, Sundong.

Abstract

The prior knowledge required to solve ARC is diverse. Therefore, for transformer-based models to learn this, such knowledge must be provided in the form of inductive biases. LatFormer is a model that has learned ARC problems by incorporating prior knowledge about 'grid transformations' as inductive biases into the transformer model. In this paper, we developed and introduced color attention into LatFormer, which allows the model to recognize color transformations by providing color information as an inductive bias. Color attention works by attending to the input and its colors before the masked self-attention occurs, calculating to what extent the color transformation will be reflected. An experiment was conducted to train and compare the performance of both the original LatFormer and the LatFormer with integrated color attention, using a dataset augmented according to the rules of color-related ARC problems.

Contents

Abstract i
Contents ii

Chapter 1. Introduction **1**

Chapter 2. Background and Related work **2**
2.1 LatFormer Architecture 2
2.1.1 Infusing Grid Transformation Prior Knowledge with Masked Self-Attention 2
2.1.2 The Mask Composition Method of the Lattice Mask Expert 3

Chapter 3. Methodology **4**
3.1 Color Attention 4
3.1.1 The Reason for Introducing Color Attention 4
3.1.2 The Working Mechanism of Color Attention 5

Chapter 4. Experiment **6**
4.1 Experimental Setup 6
4.2 Evaluation Metrics 7
4.3 Experimental Analysis 7

Chapter 5. Conclusions and Forthcoming Research **8**

References **9**

Summary (한글 요약문) **10**

Acknowledgments (감사의 글) **11**

Curriculum Vitae (약력) **12**

Chapter 1. Introduction

ARC (Abstract and Reasoning Corpus)[1] is a dataset designed by Francois Chollet to measure the generalization ability of artificial intelligence. In this dataset, each problem has a specific rule between the input and output [Figure 1.1]. Solving a problem in the ARC dataset means using the rule apparent in the task example pairs of that problem to derive an appropriate output from the input presented in the test. Since each problem uses different rules, solving all problems in ARC requires various categories of problem-solving abilities. Additionally, the ARC dataset provides about two to five task examples per problem. Therefore, models that need a sufficient amount of data during the training process cannot solve all the problems in the ARC dataset.

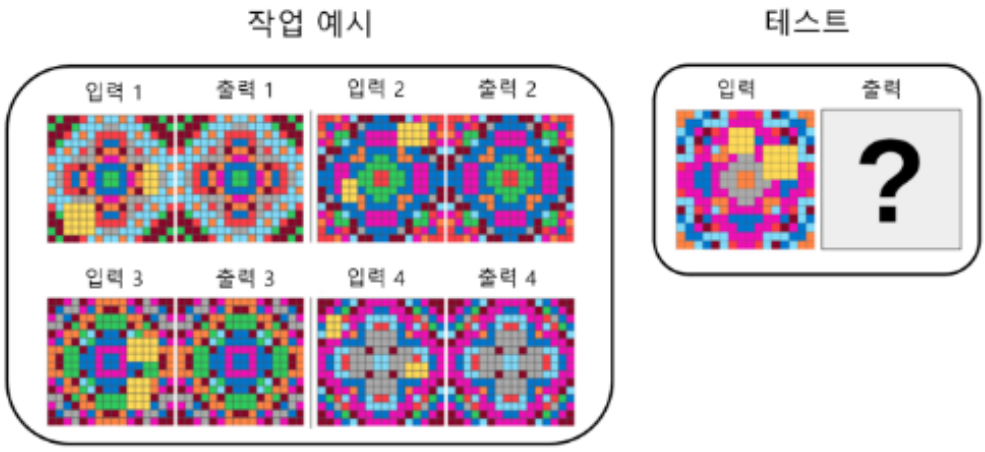

Figure 1.1: An example of an ARC problem. ARC is a dataset where the rule is inferred based on the input and output of task examples, and then the output for the input of the test is predicted.

LatFormer[2] is a model proposed for solving ARC problems, providing an inductive bias for a specific category of domain-specific language called 'grid transformations' during the training process of Transformers[3]. LatFormer utilizes domain-specific languages like movement, rotation, and flipping. However, to fully learn the ARC dataset, additional inductive biases that need to be obtained from outside the dataset are necessary.

In this paper, we introduce color attention into LatFormer, providing color information as an inductive bias to the model. Color attention is applied to pixel locations requiring color transformation, providing an inductive bias. This research is significant as it goes beyond the conventional Transformer training methods[4], which rely solely on basic color information and prompts, by introducing color attention into the Transformer model. This method of color attention can also be applied to tasks performed in computer vision.

Chapter 2. Background and Related work

2.1 LatFormer Architecture

2.1.1 Infusing Grid Transformation Prior Knowledge with Masked Self-Attention

For instance, on a Cartesian coordinate plane, if we consider four integer grid points (1, 1), (3, 2), (2, 3), and (5, 5), moving these points simultaneously by 3 units in both x and y directions results in new coordinates (4, 4), (6, 5), (5, 6), and (8, 8). Similarly, rotating the grid points by multiples of 90 degrees or mirroring them along the x or y axes also results in grid points that are integer coordinates. Such movements, rotations, and inversions, which always result in integer coordinates for the grid points, are referred to as 'grid transformations'.

ARC can be seen as a problem of inferring colors painted on grid points of a square grid, with a maximum size of 30x30. Among these, there are ARC problems that require the use of grid transformations, such as the rotation and inversion shown in [Figure 1.1]. Therefore, using grid transformations as inductive biases in model design can solve problems like those shown in [Figure 1.1]. LatFormer[2] utilizes these grid transformations as inductive biases for masked self-attention. Each mask is generated according to the input of the Transformer block, and this will be further explained in section 2.1.2. The outputted masks represent grid transformations such as movement, rotation, and inversion.

To understand the process of applying these masks, let's examine Scaled dot-product Attention[3].

$$MaskAtt(Q, K, V; M) = softmax(\frac{QK^T}{\sqrt{d}} + M)V \tag{2.1}$$

In this context, when using M as the output for the mask, the presence of - makes it non-optimizable via backpropagation. Instead, to make M differentiable, the following formula is used, applying element-wise multiplication to the matrix after softmax, thereby allowing masking:

$$MaskAtt(Q, K, V; M) = scale(softmax(\frac{QK^T}{\sqrt{d}}) \odot M)V \tag{2.2}$$

Here, ⦿ denotes element-wise multiplication, M is a matrix with values between 0 and 1, and scale() is a scaling function that adjusts the sum of each row altered due to masking back to 1. To facilitate understanding, let's assume that Q=K=V=X are column vectors and that each row of M contains only one 1 with the rest being 0s. Substituting and simplifying the expression, we get:

$$MaskAtt(X; M) = MX \tag{2.3}$$

If M represents matrices for movement, inversion, and rotation, then the result of the masked self-attention will be the vector X transformed by these operations. An example of M transforming a 5x5 input grid according to grid transformations is illustrated in [Figure 2.1]. The 5x5 grid is converted into a 25x1 vector and then multiplied by the mask from [Figure 2.1] as equation (2.3).

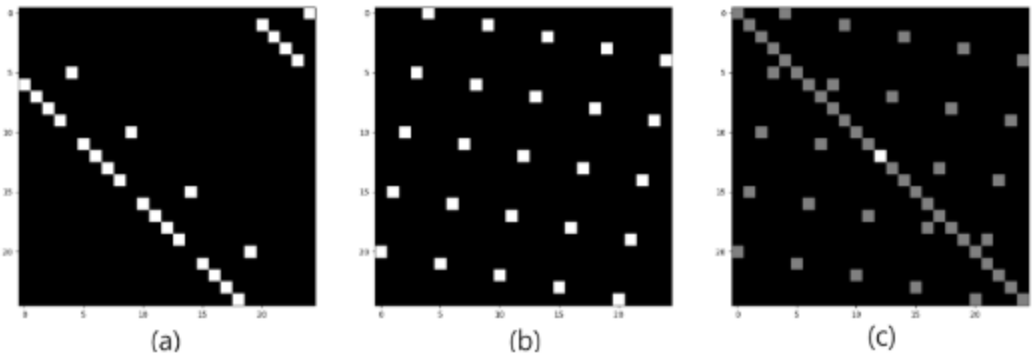

Figure 2.1: Example masks for rotating a 5x5 grid. The colors represent different values: white for 1, grey for 0.5, and black for 0. In each case, the masks illustrate (a) a one-step horizontal and vertical movement, (b) a 90-degree counterclockwise rotation, and (c) a mixture of 0-degree and 90-degree rotations.

2.1.2 The Mask Composition Method of the Lattice Mask Expert

In LatFormer, the masks are generated by a Lattice Mask Expert present in each Transformer block. The role of the Expert is to receive input data from each block and create masks for use in masked self-attention. The outputted masks, as shown in [Figure 2.1], contain detailed information about the types of grid transformations (movement, rotation, inversion) needed to solve the given problem, including the displacement of movement, the number of rotations, and the direction of inversion. The detailed information about the grid transformations is determined through the following process.

$$M_{t+1} = \alpha f(M_t) + (1 - \alpha)M_t \tag{2.4}$$

M_t represents the previous mask, f(M_t) is the mask after applying a specific grid transformation, and M_{t+1} is the output mask. The real number α, ranging between 0 and 1, is used to control the degree of mixing of grid transformations and is output through a feedforward neural network that receives input from each block. For example, let's assume f() adds a 90-degree rotation grid transformation to the mask. If M_t represents a mask that performs a 0-degree rotation, then f(M_t) will be a mask for a 90-degree rotation, and M_{t+1} will be a mask that mixes 0-degree (M_t) and 90-degree (f(M_t)) rotations based on α. By repeating this for M_{t+2}, M_{t+3}, etc., masks for 0 to 3 rotations can be generated. In this case, the feedforward neural network outputs a total of three α values to determine the required number of rotations for the input matrix. This can be applied to movement and inversion as well, and by mixing masks using each symmetry element in a weighted sum manner similar to equation (2.4), the mask for masked self-attention is completed. In actual implementation, α is output as a real number between 0 and 1, allowing for a mixture of rotated and unrotated results, as seen in [Figure 2.1] (c).

Chapter 3. Methodology

3.1 Color Attention

3.1.1 The Reason for Introducing Color Attention

LatFormer learns ARC task examples by providing inductive biases for grid transformations via masked self-attention. However, these inductive biases for grid transformations alone are insufficient to solve all problems in the ARC dataset. Therefore, to expand the range of problems that this model can solve, additional inductive biases, along with grid transformations, are necessary. Among these, this paper introduces color attention as a new method to provide the model with color information as an inductive bias, which is essential for solving ARC problems. It is expected that this color attention will enable the model to solve complex color-related problems that were previously unsolvable by conventional models.

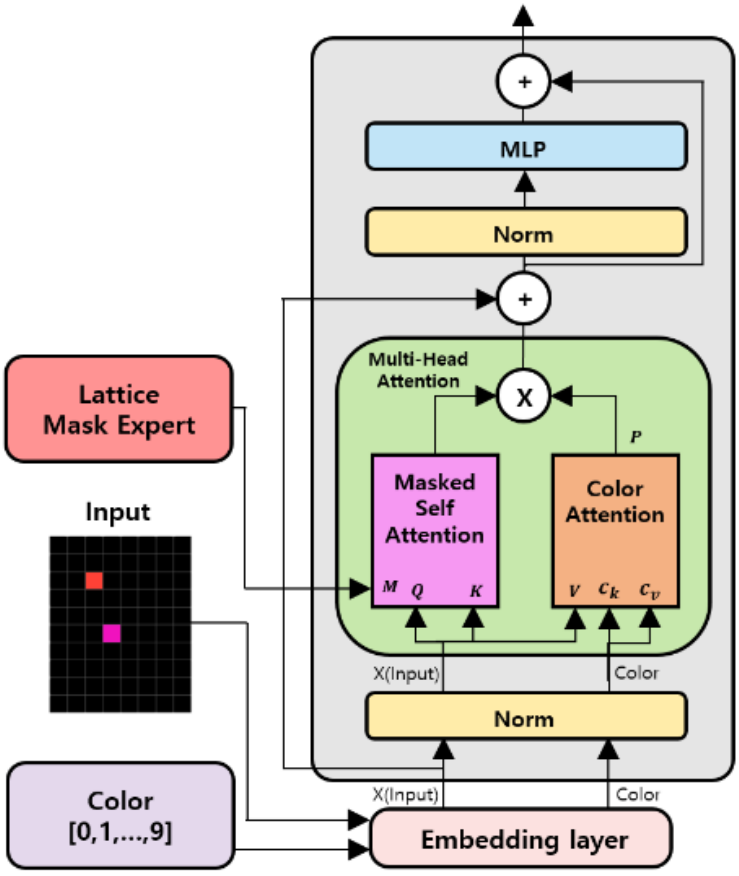

Figure 3.1: Schematic diagram of a Transformer Block with Added Color Attention

3.1.2 The Working Mechanism of Color Attention

To introduce color attention into the LatFormer model, a matrix [0, 1, . . . , 9] is used as part of the model's input along with the task example input. Let's call this matrix the color matrix. Both inputs go through the same embedding process and enter the transformer block for masked self-attention. As a result, the color matrix contains embedding information for each color based on the task example input.

Color attention occurs before V, corresponding to the task example input, is multiplied with masked self-attention. The formula for color attention is as follows, where C_k and C_v represent the transformed values of the color matrix for attention purposes.

$$ColorAttention(V, C_k, C_v) = softmax(V{C_k}^T)C_v \tag{3.1}$$

The result of this formula is the attention value for each color from 0 to 9 at each pixel of V. Similar to self-attention, this process is learned, recognizing which color is closest to the correct answer for each pixel of V and assigning weights accordingly. Then, to induce color change in the input of the task example, P is calculated by multiplying V and the result of the color attention with weights β and 1-β, respectively, and then adding them. When the result of color attention is denoted as CA, P is as follows.

$$P = (\beta V + (1 - \beta)CA) \tag{3.2}$$

β is a hyperparameter, which in this experiment is set to 0.9. The final output of the multi-head attention with added color attention is as follows.

$$Output(Q, K, V; M, P) = scale(softmax(\frac{QK^T}{\sqrt{d}}) \odot M)P \tag{3.3}$$

Chapter 4. Experiment

In this paper, we compared the performance of the LatFormer model with added color attention against the original LatFormer, particularly focusing on how much better the modified model performs on complex color problems.

4.1 Experimental Setup

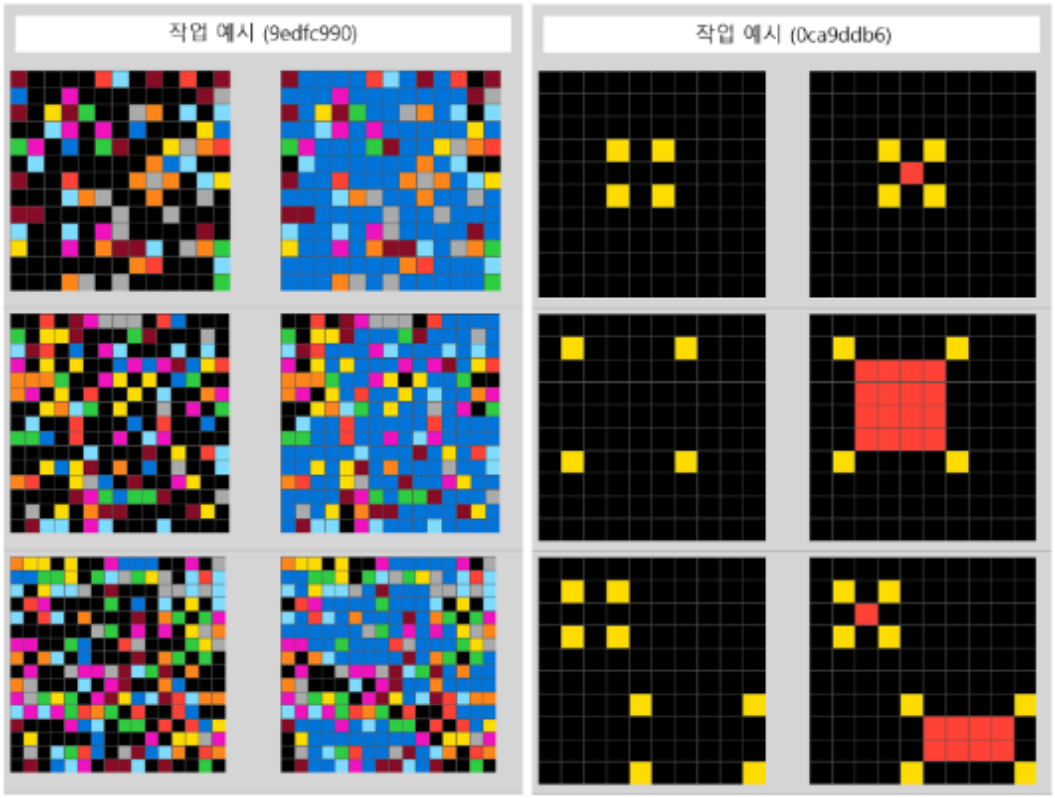

Figure 4.1: These are task examples from the ARC dataset used in the experiment. On the left is problem number 9edfc990.json, and on the right is problem number 0ca9ddb6.json.

In the experiment, the LatFormer architecture used was identical to the structure described in paper [2]. For parts of the transformer structure not additionally elaborated in the paper, the method used in ViT [5] was followed. Both the LatFormer model used in the experiment and the model with color attention converted the pixels of the task example inputs into numbers from 0 to 9 according to color before using them as inputs. The ARC problems used for the comparative experiment were 9edfc990 and 0ca9ddb6 from [Figure 4.1], which require prior knowledge about colors. To monitor the training process of both models, for each problem, 51,000 input-output pairs of size 10X10, sharing the same rule as the task example input-output, were generated. Of these, 50,000 were used as training data and 1,000 as test data. Four different random seeds were used for training the models.

4.2 Evaluation Metrics

In this experiment, Cross Entropy was used as the loss function for each of the 10X10 pixels. As a performance metric, accuracy was used, where a predicted 10X10 pixel set was considered correct if it perfectly matched the answer.

4.3 Experimental Analysis

The experimental results can be seen in [Table 1], where the results of experiments with each problem's dataset were analyzed.

For the dataset of 0ca9ddb6, both the LatFormer with color attention and the original LatFormer achieved 100% accuracy at the end of training, regardless of the random seed. This suggests that for problems utilizing simple color information, if sufficient data is provided for the training of transformer models, effective learning is possible even without additional biases.

In the dataset for 9edfc990, the average performance of the LatFormer with added color attention was 85.77%, with a standard deviation of 9.04. In contrast, the average performance of the original LatFormer was 88.56%, with a standard deviation of 5.75. When setting the null hypothesis that the model with added color attention performs better than the original LatFormer, the p-value was calculated to be 0.70. Therefore, at a 95% confidence level, the null hypothesis is rejected, indicating that there is insufficient evidence to conclude that the addition of color attention improves the learning performance of LatFormer.

Table 4.1: The table represents the accuracy at the end of training for both LatFormer and LatFormer with color attention, within a 95% confidence interval. For the task 0ca9ddb6.json, both models showed a performance of 100%. For the task 9edfc990.json, LatFormer showed an approximate performance of 88.56% ± 5.63%, whereas the LatFormer with added color attention showed about 85.73% ± 8.87%.

	LatFormer	LatFormer + 색채 어텐션
0ca9ddb6.json	100%	100%
9edfc990.json	88.56±5.63%	85.73±8.87%

Chapter 5. Conclusions and Forthcoming Research

This paper proposed the introduction of color attention to expand the range of ARC dataset problems that LatFormer can solve, enabling it to learn prior knowledge about color information. However, it was not possible to confirm whether the addition of color attention to LatFormer resulted in a significant performance improvement over the original LatFormer. Future research will likely include additional experiments to demonstrate the utility of color attention. It is also anticipated that development will focus on models that can learn multiple types of prior knowledge simultaneously, by grouping together problems that share the same prior knowledge for the model to learn.

References

1. Francois Chollet, “On the Measure of Intelligence”, arXiv:1911.01547, 2019.

2. Mattia Atzeni et al., Infusing Lattice Symmetry Priors in Attention Mechanisms for Sample-Efficient Abstract Geometric Reasoning, International Conference on Machine Learning, 2023

3. Vaswani, Ashish et al. Attention is all you need. Advances in Neural Information Processing Systems, 2017

4. Zheng, Zangwei et al. Prompt vision transformer for domain generalization. arXiv preprint arXiv:2208.08914, 2022

5. Alexey Dosovitskiy, undefined., et al, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, in International Conference on Learning Representations, 2021

Summary

Inductive bias provision for color attention learning

ARC를 풀기 위해 필요한 사전 지식은 다양하다. 그렇기에 트랜스포머 계열 모델이 이를 학습하기 위해서는 사전 지식이 귀납 편향의 형태로 제공되어야 한다. LatFormer는 '그리드 변환'에 대한 사전 지식을 트랜스포머 모델에 귀납 편향으로 넣어 ARC 문제를 학습한 모델이다. 본 논문에서는 LatFormer에 색채 정보를 귀납 편향으로 제공해 색채 변환을 인식할 수 있게 하는 색채 어텐션을 개발해 도입했다. 색채 어텐션은 마스크 자기-어텐션이 이루어지기 전 입력과 색채를 어텐션하여 색채 변환을 어느 정도 반영할 것인지 계산하는 방식으로 작동한다. 원본 LatFormer와 색채 어텐션이 포함된 LatFormer를 대상으로 색채에 관한 ARC 문제를 규칙에 맞게 증강된 데이터셋을 학습하고 성능을 비교하는 실험을 수행했다.

감 사 의 글

논문 작성을 처음 접해서 방황하고 헤매던 제게 조언을 아끼지 않으신 김선동 지도교수님께 감사드립니다. 또한 귀한 시간을 내어 논문 심사를 봐주신 안창욱 교수님께 감사드립니다. 이 외에도 제게 도움을 주신 모든 분들께 감사드립니다.

약 력

이　　름: 박지원
생 년 월 일: 2001년 11월 2일
출 생 지: 수원시
주　　소: 충청남도 서산시 읍내동 안견로 365-10

학　　력

2017. 3. – 2020. 2.　공주대학교사범대학부설고등학교
2020. 2. – 2024. 2.　광주과학기술원 전기전자컴퓨터공학부 (학사)