Chess Game Concepts Emerge under Weak Supervision: A Case Study of Tic-tac-toe

Hao Zhao[*]**& Ming Lu**
Department of Electronic Engineering
Tsinghua University
Beijing, China
{zhao-h13,lu-m13}@mails.tsinghua.edu.cn

Anbang Yao & Yurong Chen
Cognitive Computing Laboratory
Intel Labs China
Beijing, China
{anbang.yao,yurong.chen}@intel.com

Li Zhang
Department of Electronic Engineering
Tsinghua University
Beijing, China
{chinazhangli}@mail.tsinghua.edu.cn

Abstract

This paper explores the possibility of learning chess game concepts under weak supervision with convolutional neural networks, which is a topic that has not been visited to the best of our knowledge. We put this task in three different backgrounds: (1) deep reinforcement learning has shown an amazing capability to learn a mapping from visual inputs to most rewarding actions, without knowing the concepts of a video game. But how could we confirm that the network understands these concepts or it just does not? (2) cross-modal supervision for visual representation learning has drawn much attention recently. Is this methodology still applicable when it comes to the domain of game concepts and actions? (3) class activation mapping is widely recognized as a visualization technique to help us understand what a network has learnt. Is it possible for it to activate at non-salient regions? With the simplest chess game tic-tac-toe, we report interesting results as answers to those three questions mentioned above. All codes, pre-processed datasets and pre-trained models will be released.

1 Introduction

1.1 Application Background

Deep reinforcement learning (DRL) has drawn quite much attention since the publication of influential work Mnih et al. (2015). A convolutional neural network (CNN) is used to bridge the gap between video game screen frames and the most rewarding actions. An amazing feature of this kind of systems is that they do not need to know the concepts of these games (e.g. DRL learns to play *Breakout* without knowing there is a paddle or a ball in Fig 1a). However, how could we confirm that this network really understands these concepts or it just learns a mapping from patterns in the visual inputs to the best actions? This is the first question we are trying to answer here.

Mnih et al. (2015) provides some unsupervised analysis results for visualization, showing that perceptually dissimilar frames may produce close rewards, yet this does not answer the question. We choose another visualization technique called *class activation mapping* as described in Zhou et al. (2016), which can reveal where the CNN's attention is. However, directly applying it in tasks like *Breakout* still cannot answer the question. Imagine one modifies the network described in Mnih et al. (2015) into another version as Zhou et al. (2016) does. The CNN's attention may be fixed on the ball but it is still not enough to support that the network understands the concept of a ball.

---

[*]This work was done when Hao Zhao was an intern at Intel Labs China, supervised by Anbang Yao who is responsible for correspondence.
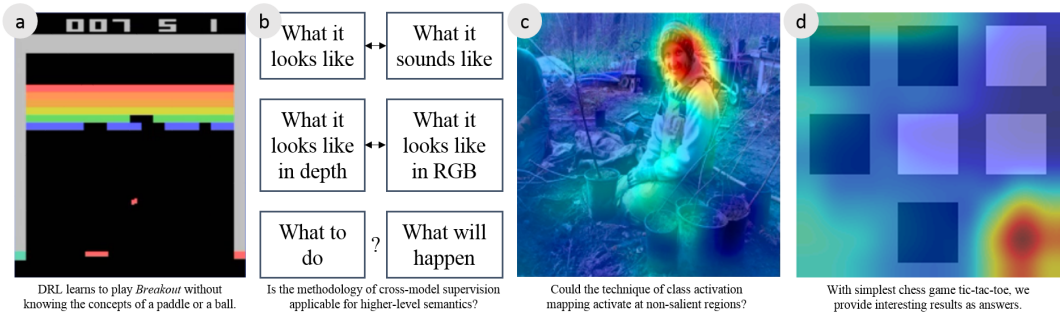
Figure 1: We raise three questions from application, methodology and technique perspectives respectively and provide our answers with a case study of the simplest chess game tic-tac-toe.

We propose to use a simple chess game called tic-tac-toe for case study. In order to answer the question, we propose a protocol as this: to place a piece where the CNN's attention is, and examine whether it is the right move. Of course, the training has to be done under weak supervision, or say, without telling the network what exactly a right move is. We think if this experiment succeeds we can claim that the network figures out the concepts of: (1) a chess board grid; (2) the winning rule; (3) two sides. Detailed analysis about these three concepts are provided later.

1.2 METHODOLOGY BACKGROUND

There have been some works about representation learning with cross-modal supervision recently. Owens et al. (2016) clusters sound statistics into several categories, and uses them as labels to learn visual representation from images corresponding to these sounds. It quantitatively shows that visual representation learnt in this way is capable of handling challenging computer vision tasks and qualitatively shows that visual and sound representations are consistent (e.g. babies' faces correspond to baby cry sound samples). Castrejón et al. (2016) goes even further by learning representations across five modalities: RGB images, clip art pictures, sketches, texts and spatial texts. Gupta et al. (2016) learns depth image representation with mid-level features extracted from RGB images as supervision, and reports improved RGB-D object detection performance.

What is the common point among these works? They generate weak supervision from one modality and use it to learn representation from another (e.g. to learn *what a train looks like* from *what a train sounds like* or to learn *what a chair looks like in depth images* from *what a chair looks like in RGB images*). During training phase, no concepts about a *train* or a *chair* are explicitly modeled. Although there are many other modalities not visited by this methodology, we think the basic ideas behind these works are same: an abstract concept like *a train* can be observed in different modalities and different representations can be connected.

Here comes the question: is this methodology still applicable when it goes beyond the problem of learning representations from different observations of a same concept? Albanie & Vedaldi (2016) is an example, which tries to relate facial expressions with what happened in a TV show (e.g. if a character earns a lot of money, she will be very happy). Although in Albanie & Vedaldi (2016) *what happened* is explicitly defined, it still can be regarded as a weak supervision for *what this expression is*.

Although with the same methodology, the problem studied in this paper addresses even higher semantics: to learn *what to do* under the weak supervision of *what will happen* (Fig 1b). This is substantially different from cross-modal supervision works mentioned above because there is no longer a certain abstract concept of object or attribute observed in different modalities. Instead, figuring out the relationship between *what to do* and *what will happen* needs a higher level of intelligence.

1.3 TECHNIQUE BACKGROUND

The core technique used in this paper is *class activation mapping (CAM)* as described in Zhou et al. (2016). So leaving out all the backgrounds about playing a chess game or cross-modal supervision, what do our experiments say more than its inventors'? We think we show that CAM can also activate at non-salient regions. CAM helps us to understand where contributes the most to a classification result. As Fig 1c shows, the heatmap reveals that the face contributes the most to the result that the network claims it as a *person*.

As has already been shown by Krizhevsky et al. (2012), kernels of lower layers of a CNN capture gradients in an image. Existing CAM experiments tend to activate at salient regions, and this is very reasonable because there are more gradients and therefore more information (e.g. the face in Fig 1c). Here comes the question: could CAM activate at non-salient regions like the empty spaces on a chess board? Our answer is positive as the results (Fig 1d) show that in order to predict *what will happen* in the future, the CNN's attention is fixed upon texture-free regions.

Since we render chessboards as visual inputs without adding noise, those *empty* spaces are completely *empty* meaning that: (1) if we take out the activated patch in Fig 1d, all pixels in this patch have exactly the same value. (2) If we evaluate this patch with quantitative information metric like entropy, there is no information here. Thus the only reason why these regions are activated is that the network collects enough information from these regions' receptive fields. We argue that this experiment (CAM can activate at non-salient regions) testifies (again) CNN's ability to hierarchically collect information from visual inputs.

1.4 WHAT THIS PAPER IS ABOUT

After introducing those three backgrounds, we describe our work briefly as: to classify rendered tic-tac-toe chessboards with weak labels and to visualize that the CNN's attention automatically reveals where the next piece should be placed. Learnt representation shows that: (1) the network knows some concepts of the game that it is not told of; (2) this level of supervision for representation learning is possible; (3) the technique of class activation mapping can activate at non-salient regions.

2 RELATED WORKS

2.1 CONCEPT LEARNING

Concept learning has different meanings in different contexts, and how to confirm a concept is learnt remains an open question. In Jia et al. (2013), a *concept* is learnt if a generative model is learnt from a small number of positive samples. In Lake et al. (2015), a *concept* is learnt if a model learnt from only one instance can generalize to various tasks. Higgins et al. (2016) claims a *concept* is learnt when a model can predict unseen objects' sizes and positions. To summarize, they evaluate whether a concept is learnt through a model's generalization ability. In even earlier works like Zhu et al. (2010);Yang et al. (2010), concept learning means a object/attribute classification task dealing with appearance variations, in which a concept is actually already pre-defined.

Unlike these works, we investigate the concepts of game rules instead of object/attribute. Unlike Jia et al. (2013);Lake et al. (2015);Higgins et al. (2016), we claim a *concept* is learnt through a novel testing protocol instead of generalization ability. Why generalization ability could show a concept is learnt? We think the reason is that a model understands a *concept* if it can use it in more cases. To this end, we argue that our protocol could also show a *concept* is learnt because the learnt representations in our experiments can be used to decide *what to do* though no rule about what need to be done is provided.

2.2 CROSS-MODAL SUPERVISION

The literature of cross-model supervision and the differences between this paper and existing ones are already covered in last section. Here we re-claim it briefly: Owens et al. (2016);Castrejón et al. (2016);Gupta et al. (2016) learn representations across modalities because actually they are different observations of a same (object or attribute) concept. Whether this methodology is applicable for
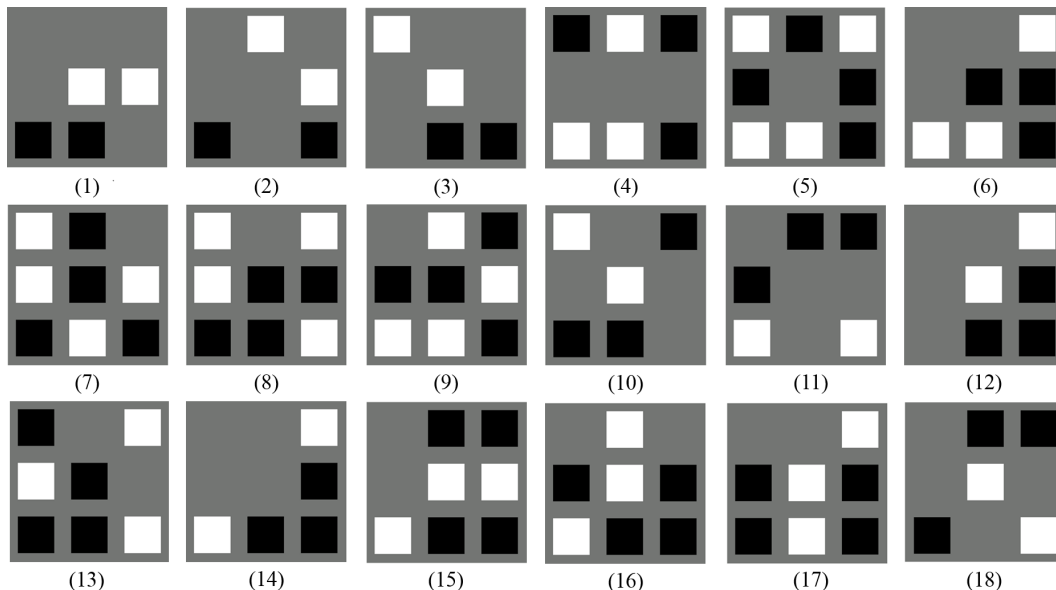
Figure 2: 18 different types of chessboard states and corresponding labels.

higher-level concepts like game rules remains an open question and we provide positive answers to this question.

2.3 CLASS ACTIVATION MAPPING

Before the technique of class activation mapping is introduced by Zhou et al. (2016), pioneering works like Simonyan et al. (2014);Zhou et al. (2015) have already shown CNN's ability to localize objects with image-level labels. Although with different techniques, Simonyan et al. (2014);Zhou et al. (2015)'s activation visualization results also focus on salient regions. Unlike these works, we show that class activation mapping can activate at non-salient regions, or say more specifically, completely texture-free regions. Since the activated patch itself provides no information, all discriminative information comes from its context. This is another strong evidence to prove CNN's capability to collect information from receptive fields, as a hierarchical visual model.

3 EXPERIMENT I: GAME ENDS IN NEXT MOVE

A tic-tac-toe chessboard is a 3×3 grid, and there are two players (black and white in our case). Due to duality, we generate all training samples assuming the black side takes the first move. The state space of tic-tac-toe is small consisting of totally $3^9 = 19683$ combinations. Among them, many combinations are illegal such as the one in which all 9 pieces are black. We exhaustively search over the space according to a recursive simulation algorithm, in which: (1) the chessboard state is denoted by an integer smaller than 19683. (2) every state corresponds to a 9-d vector, with each element can take a value from this set {0-illegal, 1-black win, 2-white win, 4-tie, 5-uncertain}. We call this 9-d vector a state transfer vector, denoting what will happen if the next legal piece placement happens at according location. (3) generated transfer vectors can predict the existence of a critical move that will finish the game in advance. We will release this simulation code.

After pruning out illegal states, we collect 4486 possible states in total. Among these samples, we further take out 1029 states that a certain side is going to win in the next move. We then transform these chessboard states into visual representations (gray-scale images at resolution (180, 180)). Each of these 1029 samples is assigned a label according to the state transfer vectors. There are totally 18 different labels illustrating 2 (sides) $\times$ 9 (locations). As demonstrated by Fig 2, we randomly pick a sample for each label. As mentioned before black side takes the first move, thus if the numbers of
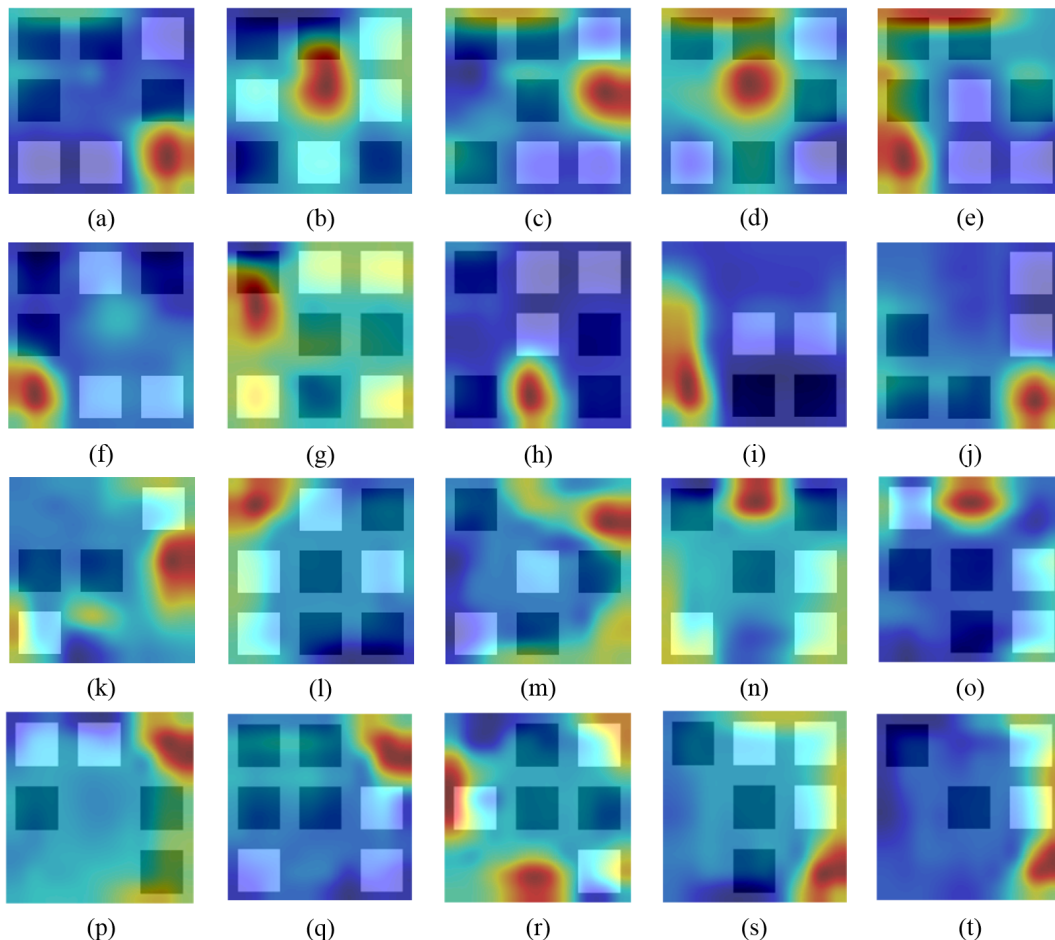
Figure 3: Class activation mapping results on our dataset.

black and white pieces are equal the next move will be black side's and if there are one more black piece the next move will be white side's.

Although the concepts of two sides and nine locations are coded into the labels, this kind of supervision is still weak supervision. Because what we are showing to the algorithm is just 18 abstract categories as Fig 2 shows. Could an algorithm figure out what it needs to do by observing these visual inputs? We think even for a human baby it is difficult because no concepts like *this is a game* or *you need to find out how to win* are provided. In the setting of deep reinforcement learning there is at least an objective of *getting higher score* to pursue.

As mentioned before, the method we exploit is to train a classification network on this rendered dataset (Fig 2) and analyze learnt representations with the technique of class activation mapping. As Zhou et al. (2016) suggests, we add one global average pooling layer after the last convolutional layer of a pre-trained AlexNet model. All fully connected layers of the AlexNet model are discarded, and a new fully connected layer is added after the global average pooling layer. After the new classification network is fine-tuned on our dataset, a CAM visualization is generated by weighting the outputs of the last convolutional layer with parameters from the added fully connected layer. Our CAM implementation is built upon Marvin and it will be released.

Due to the simplicity of this classification task, the top one classification accuracy is 100% (not surprisingly). Class activation mapping results are provided in Fig 3 and here we present the reasons why we claim concepts are learnt: (1) We provide 18 abstract categories, but in order to classify visual inputs into these 18 categories the network's attention is roughly fixed upon chessboard grids.
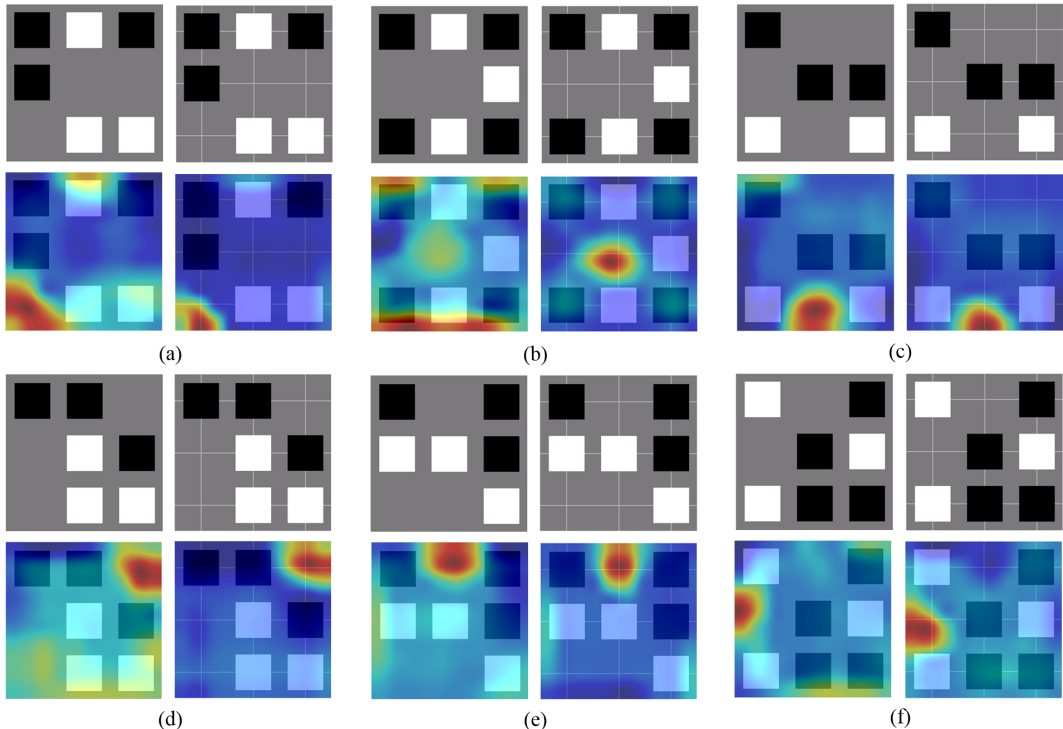
Figure 4: Class activation mapping results after grid lines are added.

This means the concept of grid emerges in the learnt representation. (2) If we place a piece at the most activated location in Fig 3, that will be the right (and legal) move to finish the game. On one hand, this means the concept of winning rule emerges in the learnt representation. On the other hand, this means this learnt concept can be used to deal with un-taught task (analogous to Jia et al. (2013);Lake et al. (2015);Higgins et al. (2016) who use generalization ability to illustrate that concepts are learnt). (3) As Fig 3cehijnpq show, both sides can win in the next move if we violate the take-turns rule. However, the network pays attention to the right location that is consistent to the rule. For example, in Fig 3j, it seems that placing a black piece at the left-top location will also end the game. However, this move will violate the rule because there are already more black pieces than white pieces meaning that this is the white side's turn. This means that the concept of two sides emerges in learnt representation.

Except for learnt concepts, we analyze what this experiment provides for the remaining two questions. To the second question: results in Fig 3 show that the methodology of generating labels from one modality (state transfer vectors in our case) to supervise another modality is still applicable. More importantly, we use images as inputs yet the learnt visual representations contain not only visual saliency information but also untold chess game concepts. To the third question: as Fig 3 shows, most activated regions are empty spaces on the chessboard.

4 EXPERIMENT II: ADDING GRID LINES

Since we claim complicated concepts emerge in learnt visual representations, a natural question will be: if the chessboard's and pieces' appearances are changed does this experiment still work? Thus we design this experiment by adding grid lines to the chessboards when rendering synthetic data (Fig 4). The intentions behind this design is three-folded: (1) in this case, the chessboard's appearance is changed. (2) after these lines are added, the concept that there is a chessboard grid is actually implied. Still, we do not think these lines directly *provide* the concept of chessboard grid thus we use the word *imply*. Whether the network can figure out what these lines mean still remain
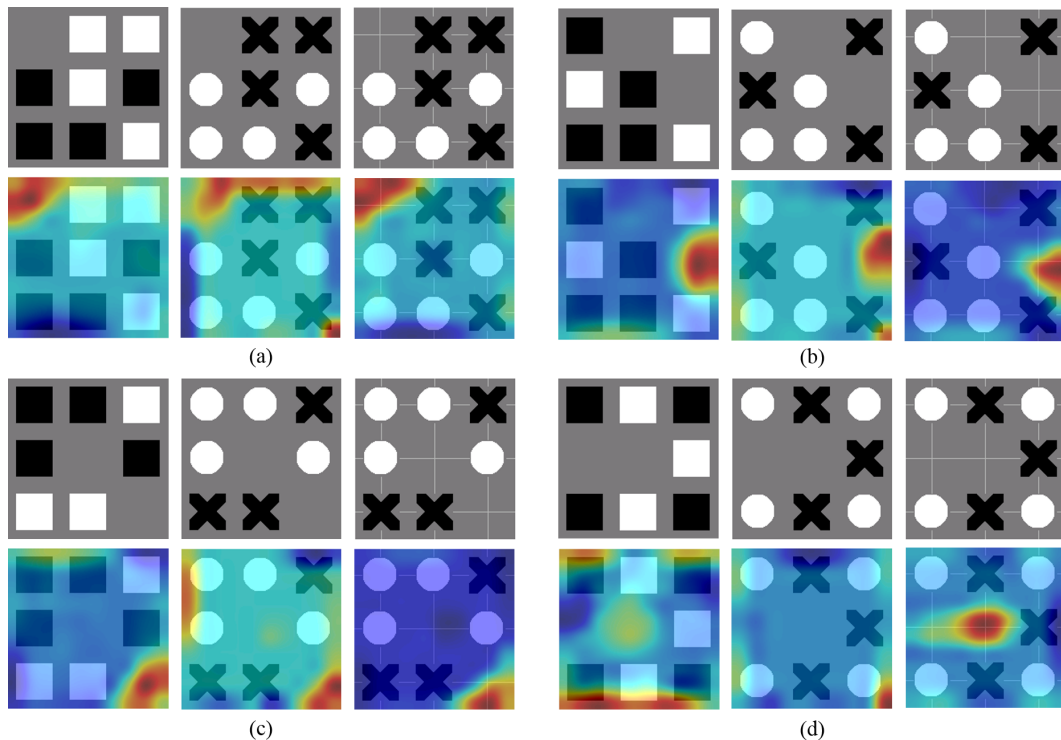
Figure 5: Class activation mapping results after piece appearance is changed.

uncertain. (3) those locations that are *completely* empty in Experiment I are no longer empty from the perspective of information (still empty from the perspective of game rule).

We train the same network on the newly rendered dataset with grid lines and calculate CAM results in the same way. The results are demonstrated by Fig 4. Generally speaking, the grid lines allow the network to better activate at the location of right move, making them stands out more on the heatmap. What does this mean to the three intentions mentioned in last paragraph? (1) Firstly, it shows that our experiment is robust to chess board appearance variance. (2) Secondly, after implying the concept that there is a chessboard grid, the network performs better at paying attention to the location of right move. Again we compare this phenomenon against how a human baby learns. Although not supported by phycological experiment, we think with a chessboard grid a human baby is more easy to figure out the game rule than without. (3) Thirdly, heatmap changes in Fig 4 is not surprising, because after adding those lines, the *empty* (from the perspective of game rule) regions contain more gradients for lower layers of a CNN to collect. However, again it supports that *activating at non-salient regions* is *NOT* trivial.

5 EXPERIMENT III: PIECE APPEARANCE CHANGE

In this experiment we change the appearance of the piece by: (1) replacing black boxes with white circles; (2) replacing white boxes with black crosses. Note that in this case the white side moves first. Again we train the same network and visualize with CAM. The results comparison is provided in Fig 6. Further we add grid lines to the cross/circle chessboard.

6 EXPERIMENT IV: MODEL BEHAVIOR OVER TIME

In order to further demonstrate the non-triviality of the model behaviors, we design this experiment. We train on the dataset in Experiment I with 1000 iterations and snap-shotted the parameters at 500th iteration. The classification accuracy is 100% at 1000th iteration and 53.13% at 500th iteration. The
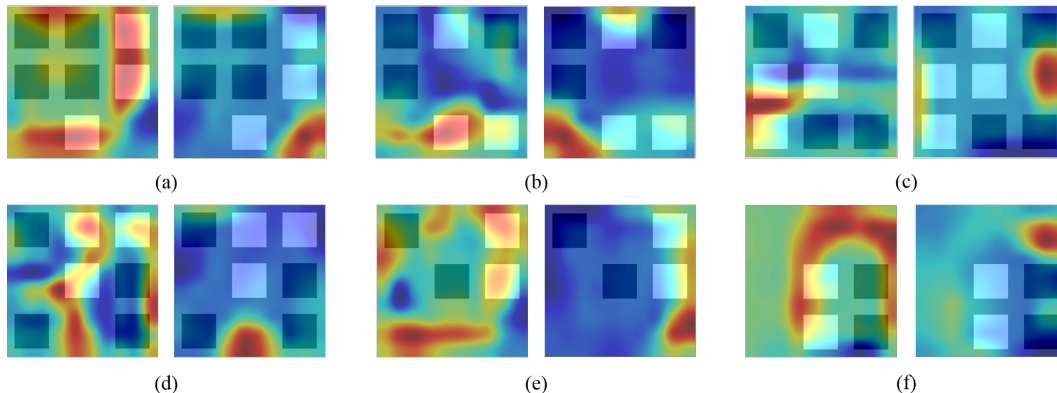
Figure 6: Class activation mapping results on true positive samples at 500 iterations (left, 53.13% accuracy) and 1000 iterations (right, 100% accuracy).
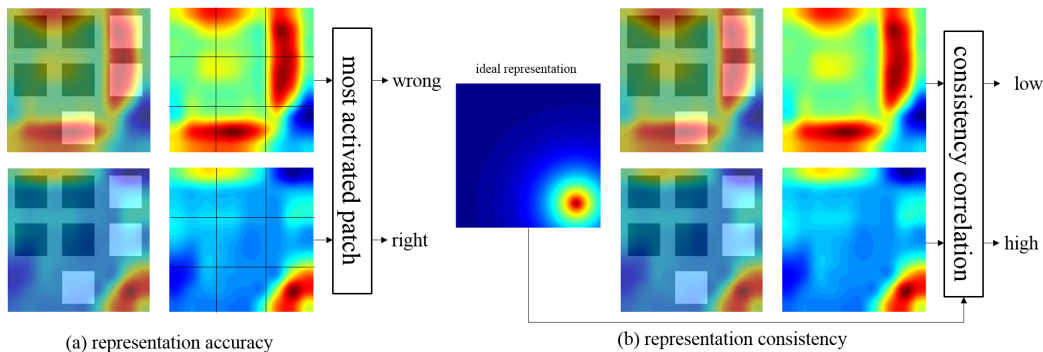

Figure 7: We propose two quantitative evaluation protocols: (a) by selecting the most activated patch, we calculate how frequent the representation fire at the correct location; (b) we correlate the representation with an ideal activation map.

CAM results are shown by Fig 5 in which all samples are true positives. We think it shows that there are two ways to achieve this classification task: (1) by paying attention to the visual patterns formed by the existing pieces; (2) by paying attention to where the next piece should be placed. This experiment shows that at an earlier stage of learning the model's behavior is consistent to the first hypothesis and after the training is completely done the network can finally fire at correct location.

7   QUANTITATIVE EVALUATION

We propose two different quantitative evaluation protocols. The first one is representation accuracy (RAC), for which we select the most activated patch and examine whether it is the correct location to end the game. The second one is representation consistency (RCO), which correlates the normalized representation and a normalized ideal activation map. The quantitative comparisons are shown in Table 1, in which NAC stands for network classification accuracy. These results quantitatively support that: (1) learnt representation can be used to predict the right move at an over 70% accuracy. (2) adding grid lines (implying the concept of a chessboard) dramatically improves localization.

8   CONCLUSION

The core experiment in this paper is to train a classification CNN on rendered chessboard images under weak labels. After class activation mapping visualization, we analyse and interpret the results

Experiment	I original	II grid	III piece	III piece+grid	IV 500th
NAC (%)	100.00	100.00	100.00	100.00	53.13
RAC (%)	71.82	97.25	83.77	99.00	27.87
RCO (10^3)	-8.096	-5.115	-7.751	-4.9321	-10.610

Table 1: Quantitative results.

in three different backgrounds. Although simple, we argue that our results are enough to show that: (1) a CNN can automatically figure out complicated game rule concepts in this case. (2) cross-modal supervision for representation learning is still applicable in this case of higher-level semantics. (3) the technique of CAM can activate at non-salient regions, testifying CNN's capability to collect information from context in an extreme case (only context has information).

REFERENCES

Samuel Albanie and Andrea Vedaldi. Learning grimaces by watching tv. In *BMVC*, 2016.

Lluıs Castrejón, Yusuf Aytar, Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Learning aligned cross-modal representations from weakly aligned data. In *CVPR*, 2016.

Saurabh Gupta, Judy Hoffman, and Jitendra Malik. Cross modal distillation for supervision transfer. In *CVPR*, 2016.

Irina Higgins, Loic Matthey, Xavier Glorot, Arka Pal, Benigno Uria, Charles Blundell, Shakir Mohamed, and Alexander Lerchner. Early visual concept learning with unsupervised deep learning. *arXiv:1606.05579*, 2016.

Yangqing Jia, Joshua T Abbott, Joseph Austerweil, Thomas Griffiths, and Trevor Darrell. Visual concept learning: Combining machine vision and bayesian generalization on concept hierarchies. In *NIPS*, 2013.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. In *Science*, 2015.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. In *Nature*, 2015.

Andrew Owens, Jiajun Wu, Josh H McDermott, William T Freeman, and Antonio Torralba. Ambient sound provides supervision for visual learning. In *ECCV*, 2016.

Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. 2014.

Jingjing Yang, Yuanning Li, Yonghong Tian, Ling-Yu Duan, and Wen Gao. Per-sample multiple kernel approach for visual concept learning. In *Journal on Image and Video Processing*, 2010.

Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Object detectors emerge in deep scene cnns. In *ICLR*, 2015.

Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *CVPR*, 2016.

Shiai Zhu, Gang Wang, Chong-Wah Ngo, and Yu-Gang Jiang. On the sampling of web images for learning visual concept classifiers. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2010.