Tristan Hascoet
Kobe University
tristan@people.kobe-u.ac.jp

To-san
Kobe University
xxx

Mec people1
xxx
xxx
xxx

Mec people2
xxx
xxx
xxx

Yasuo Ariki
xxx
xxx
xxx

Tetusya Takiguchi
xxx
xxx
xxx

Abstract

xxx

1. Introduction

Deep Learning models are often viewed as black box predictors lacking interpretability in the sense that existing tools often fail to explain the decision making process behind the models predictions. In contrast, humans are often able to justify in natural language the reason behind their answer to a certain visual question. While it is true that humans can justify for their answers on high level reasoning tasks (VQA, spatial and causal relationships), humans also often fail to explain the process behind their low-level feature recognition ability: for example, precisely defining the nature of a specific texture and a low-level part attributes exhibiting large intra-class variations such as "legs" or "wings"

Oftentimes humans expert are asked to perform such low-level visual recognition tasks while being unable to justify for their recognition process. In this paper, we present one very practical instance of such situation in the microprocessor chip industry, in which expert material scientists are tasked with assessing the quality of a copper sheet surface after degradation by atmospheric conditions. We propose a model that, under careful design considerations, is able to provide visual clues for human experts to understand and justify for their own decision process.

Our model is carefully designed so that a subset of its internal representations carry semantically meaningful Information that can be visualized and easily interpreted by the human experts, confirming their intuitions and eventually shading light on their own decision processes and biases.

xxx NEED INDUCTIVE BIAS MENTION xxx

We then show how these semantically meaningful representations can be used To formulate and validate hypothesis on the physical phenomenon underlying the observed signal; i.e. the degradation of the copper sheet surface as a function of time and atmospheric conditions.

Going one step further we show how, given sufficient experimental data, the model can be augmented to automatically learn and formulate these hypothesis on itself.

In essence, the argument this paper is aiming for is as follows: although deep learning models lack the "common sense reasoning abilities and the powerful formalism of natural language to communicate and justify for their decision making process, they can provide powerful tools to shed some light into the low-level recognition process they share with humans [**?**], and for which human introspection fails to provide convincing explanations.

In practice, the contribution of this paper is as follows: - We formalize a segmentation procedure for statistical surface based on probabilistic lazy label segmentation framework - We show how model generalization can be used as a proxy metric to quantify the validity of an hypothesis. - We propose a simple model of material surface degradation through time and decompose a model architecture into a recognition module and an hypothesis module that directly learns to formulate the most pertinent hypothesis based on

The remainder of this paper is organized as follows: In Section 2, we present background information on our task: we detail the practical stakes and problem definition, etc. In Section 3, we detail our approach and etc. Section 4 relates our work to existing research Section 5 presents the experiments setting and discuss our results and Section 6 concludes this paper.

2. Background

xxx
xxx
xxx

xxx
xxx
xxx

3. Method

3.1. Framework

xxx

3.2. Recognition module

xxx

3.3. Hypothesis formulation

xxx

3.4. Hypothesis module

xxx

4. Related Work

xxx
xxx
xxx
xxx
xxx

5. Experiments

xxx

5.1. Recognition results

xxx

5.2. Hypothesis testing

xxx

5.3. Hypothesis learning

xxx

6. Conclusion

References