Deep learning based Feature Selection Study

Sanghyuck Lee

2021. 06 29.

Outline

- Deep Learning based Feature Selection study
- 중견 연구 과제
- 갑상선 연구 과제
- 산학 협력 과제

Outline of Deep Feature Selection

- Intro: Last Discussion - What is Reproducible Feature Selection
- Meaning of 'Reproducible'
- DeepPINK Review
  - Interpretability of DNN and Reproducibility of Discoveries
  - The Role of FDR on Reproducible Results (Real-Effects)
  - Summary
  - Flowchart
- Interpretation of Neural Networks Is Fragile
- Discussion

Intro: Last Discussion

- What is Reproducible Feature Selection?

Lu, Yang Young, et al. "DeepPINK: reproducible feature selection in deep neural networks." *arXiv preprint arXiv:1809.01185* (2018).

Meaning of ‘Reproducible’

The dictionary definition (Naver)

reproducible 미국식 영국식

형용사 재생[재현]할 수 있는, 복사[복제]할 수 있는, 모사[모조]할 수 있는. ~bíl·i·ty

reproducibility 미국식 영국식

명사 재생력, 재현 가능성; 복사[복제] 가능성.

The meaning in science (WiKi pedia)

Reproducibility is a major principle of the scientific method. It means that a result obtained by an experiment or observational study should be achieved again with a high degree of agreement when the study is replicated with the same methodology by different researchers. Only after one or several such successful replications should a result be recognized as scientific knowledge.

Goodman, Fanelli, & Ioannidis (p. 2)

- **Methods reproducibility**: The ability to implement, as exactly as possible, the experimental procedures, with the same data and tools, to obtain the same results.
- **Results reproducibility**: The production of corroborating results in a new study, having followed the same experimental methods.
- **Inferential reproducibility**: The making of knowledge claims of similar strength from a study replication or reanalysis.

DeepPINK: Abstract

- Abstract section in DeepPINK

"Deep learning has become increasingly popular in both supervised and unsupervised machine learning thanks to its outstanding empirical performance. However, because of their intrinsic complexity, most deep learning methods are largely treated as **black box tools with little interpretability**. Even though recent attempts have been made to facilitate the interpretability of deep neural networks (DNNs), existing methods are susceptible to noise and lack of robustness. **Therefore**, scientists are justifiably cautious about the **reproducibility of the discoveries**, which is often related to the **interpretability** of the underlying statistical **models**. In this paper, we describe a method to increase the **interpretability and reproducibility of DNNs** by incorporating the idea of feature selection with controlled error rate. By designing a new DNN architecture and integrating it with the recently proposed knockoffs framework, we perform feature selection with a controlled error rate, while maintaining high power"

DeepPINK: Abstract

<Abstract of DeepPINK>
Little interpretability of DNNs <-> Reproducibility of the discoveries
->Interpretability and reproducibility of DNNs

- A

"D chine lea sic complexity, most deep learning methods are largely treated as **black box tools with little interpretability**. Even though recent attempts have been made to facilitate the interpretability of deep neural networks (DNNs), existing methods are susceptible to noise and lack of robustness. **Therefore**, scientists are justifiably cautious about the **reproducibility of the discoveries**, which is often related to the **interpretability** of the underlying statistical **models**. In this paper, we describe a method to increase the **interpretability and reproducibility of DNNs** by incorporating the idea of feature selection with controlled error rate. By designing a new DNN architecture and integrating it with the recently proposed knockoffs framework, we perform feature selection with a controlled error rate, while maintaining high power"

DeepPINK: Interpretability of DNN and Reproducibility of Discoveries

- Part of Introduction section in DeepPINK

"… In many scientific areas, **interpretability** of scientific results is becoming increasingly important, as researchers aim to understand **why and how a machine learning system makes a certain decision**. For example, if a clinical image is predicted to be either benign or malignant, then the doctors are eager to know which parts of the image drive such a prediction. … Therefore, an explainable and interpretable system to reason about why certain decisions are made is critical to convince domain experts [27]. … Recently, identifying explainable features that contribute the most to DNN predictions has received much attention. … In such a setting, it is desirable for practitioners to **select explanatory features** in a fashion that is robust and **reproducible**, even in the presence of noise. Though considerable work has been devoted to creating feature selection methods that select relevant features, it is less common to carry out feature selection while explicitly controlling the associated error rate. Among different feature selection performance measures, the **false discovery rate (FDR) [4] can be exploited to measure the performance of feature selection algorithms**. Informally, the FDR is the expected proportion of falsely selected features among all selected features, where a **false discovery is a feature that is selected but is not truly relevant** (For a formal definition of FDR, see section 2.1). …"

DeepPINK: Interpretability of DNN and Reproducibility of Discoveries

- Part of Introduction section in DeepPINK

"… In rtant, as rese **decisi** the doctor explai al to convin most t practitioners to **select explanatory features** in a fashion that is robust and **reproducible**, even in the presence of noise. Though considerable work has been devoted to creating feature selection methods that select relevant features, it is less common to carry out feature selection while explicitly controlling the associated error rate. Among different feature selection performance measures, the **false discovery rate (FDR) [4] can be exploited to measure the performance of feature selection algorithms**. Informally, the FDR is the expected proportion of falsely selected features among all selected features, where a **false discovery is a feature that is selected but is not truly relevant** (For a formal definition of FDR, see section 2.1). …"

Identifying explainable features in DNNs -> interpretability of DNNs,
but existing works are vulnerable to some small noises.
-> Reproducible feature selection
Discovery->Selected Feature
Reproducibility of Discoveries -> Reproducibility of Selected Features

DeepPINK: The Role of FDR on Reproducible Results (Real-Effects)

Barber, Rina Foygel, and Emmanuel J. Candès. "Controlling the false discovery rate via knockoffs." *Annals of Statistics* 43.5 (2015): 2055-2085.

"… Imagine we have a procedure that has just made 100 discoveries. Then roughly speaking, if our procedure is known to control the FDR at the 10% level, this means that we can expect at most 10 of these discoveries to be false and, therefore, at least 90 to be true. In other words, if the collected data were the outcome of a scientific experiment, then we would expect that most of the **variables selected** by the knockoff procedure correspond to **real effects that could be reproduced in follow-up experiments.** … "

DeepPINK: The Role of FDR on Reproducible Results (Real-Effects)

- Part of Introduction section in DeepPINK

In this paper, we integrate the idea of a knockoff filter with DNNs to enhance the **interpretability** of the learned network model. Through simulation studies, we discover surprisingly that naively combining the knockoffs idea with a multilayer perceptron (MLP) yields extremely low power in most cases (though FDR is still controlled). To resolve this issue, we **propose a new DNN architecture named DeepPINK** (Deep feature selection using Paired-Input Nonlinear Knockoffs). DeepPINK is built upon an MLP with the major distinction that it has an additional pairwise-connected layer containing p filters, one per each input feature, where each filter connects the original feature and its knockoff counterpart. We demonstrate empirically that **DeepPINK achieves FDR control with much higher power** than many state-of-the-art methods in the literature. ….

Identifying explainable features in DNNs -> interpretability of DNNs,
but existing works are vulnerable to some small noises.
-> Reproducible feature selection
Discovery->Selected Feature
Reproducibility of Discoveries -> Reproducibility of Selected Features

DeepPINK: Summary

<Abstract of DeepPINK>
Little interpretability of DNNs <-> Reproducibility of the discoveries
->Interpretability and reproducibility of DNNs

<Introduction of DeepPINK >
Identifying explainable features in DNNs -> interpretability of DNNs,
but existing works are vulnerable to some small noises.
-> Reproducible feature selection
Discovery->Selected Feature
Reproducibility of Discoveries -> Reproducibility of Selected Features

<My summary of Abs and Intro of DeepPINK>
Interpretability of DNNs -> Existing strategies are not robust and hard to reproducible -> Reproducibility of Discoveries(Selected Features)
Reproducible feature selection that is DeepPINK

DeepPINK: Flowchart

Problem of Interest

Little Interpretability of DNNs

Approach of Existing Studies

Feature Selection (important feature identification)

Problem of Existing Studies

1) Hard to generalize
2) Sensitive to noise
(Very little noise affects feature importance)

The Motivation of Proposed Approach: Knockoff Filter

Controlling FDR to ensure a lower bound on the proportion of true effect features

Proposed Approach

Reproducible Feature Selection based on Knockoff Filter

Evaluation

1) FDR
2) Power
for Several Dataset

DeepPINK: Flowchart

Problem of Interest

Little Interpretability of DNNs

Approach of Existing Studies

Feature Selection (important feature identification)

Problem of Existing Studies

1) Hard to generalize
2) Sensitive to noise
(Very little noise affects feature importance)

The Motivation of Proposed Approach:

FDR vs Y prediction socore
(R^2, confidence score, or accuracy, …)

er bound on the
oportion of true effect
features

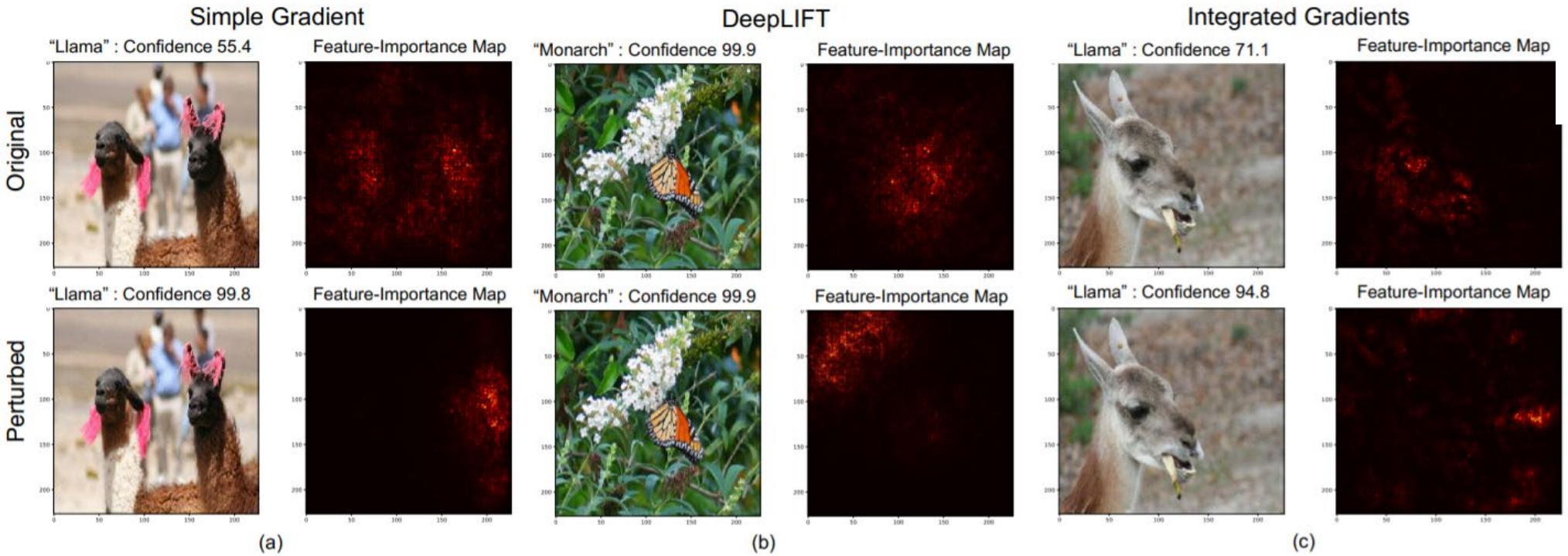
Proposed Approach

Reproducible Feature Selection based on Knockoff Filter

Evaluation

1) FDR
2) Power
for Several Dataset

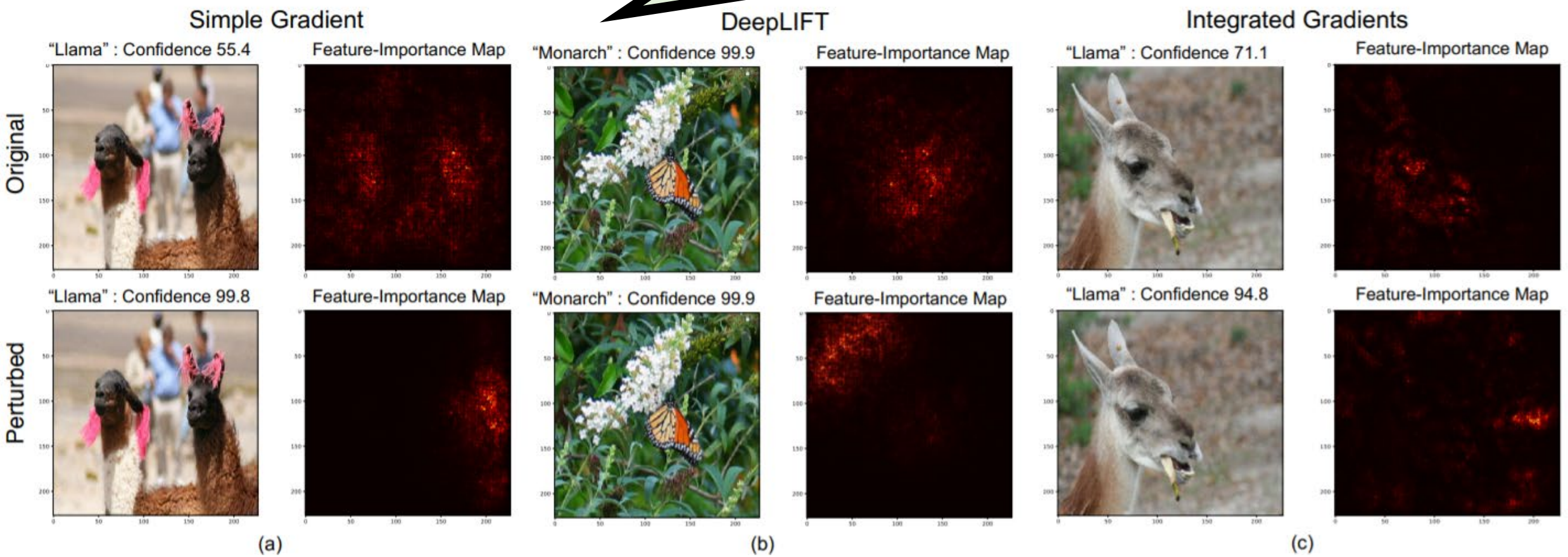
Interpretation of Neural Networks Is Fragile



Adversarial attack against feature-importance maps: We generate feature-importance scores, also called saliency maps, using three popular interpretation methods: (a) simple gradients, (b) DeepLIFT, and (c) integrated gradients. The top row shows the the original images and their saliency maps and the bottom row shows the **perturbed images** (using the center **attack** with = 8, as described in Section 2) and corresponding saliency maps. In all three images, the **predicted label does not change** from the perturbation; however, the saliency maps of the perturbed images shifts dramatically to features that would not be considered salient by human perception.

Interpretation of Neural Networks Is Fragile

FDR vs Y prediction socore
(R^2, confidence score, or accuracy, …)



Discussion

- Image Classification, Detection, Segmentation …

- FDR with Image data -> Interpretability of DNNs

중견 연구 제안서 작성

- 가제:

  한글: 의료 분야내 주요 관심 구조물 세그멘테이션을 위한 사전훈련 최신 딥러닝 모델 개발

  영문: Development of Pre-trained Deep Learning Model for Segmentation of Structures of Major Interest in the Medical Field

- 연구 목표:

  본 과제에서는 몇몇 주요한 해부학적 구조에 대해서 세계 최고 수준의 성능을 갖는 사전학습된 세그멘테이션 모델을 배포함으로써, 다양한 의료분야 응용에 성공적으로 기여하는 것을 최종 목표로 하며, 이를 위해 다음과 같은 연차별 연구 목표를 설정하여 추진한다.

- 이번 주 데이터셋 확보 회의 예정:

  정부 추진 의료 데이터셋 및 해외 오픈 데이터셋

  – 현황 조사 및 가능한 경우에 대해 실제 샘플 프레젠테이션

갑상선 연구

- 완료 현황: 각각 1237 환자

유형	눈 (L)	눈 (R)	시신경 (L)	시신경 (R)	내직근 (L)	내직근 (R)	외직근 (L)	외직근 (R)	상직근 (L)	상직근 (R)	하직근 (L)	하직근 (R)	상안검 (L)	상안검 (R)	지방 (L)	지방 (R)	눈물주머니 (L)	눈물주머니 (R)	안와 (L)	안와 (R)	평균
type	눈 (L)	눈 (R)	시신경 (L	시신경 (R	내직근 (L	내직근 (R	외직근 (L	외직근 (R	상직근 (L	상직근 (R	하직근 (L	하직근 (R	상안검 (L	상안검 (R	지방 (L)	지방 (R)	눈물 주머니 (L)	눈물 주머니 (R)	안와 (L)	안와 (R)	합계
AX 1	0.91	0.87	0.74	0.81	0.91	0.94	0.97	0.97							0.81	0.85					0.88
AX 2	0.97	0.97	0.79	0.75	0.97	0.97	0.96	0.93							0.94	0.99					0.92
AX 3	1	1	0.99	0.98	0.99	1	0.94	0.95							0.99	1					0.98
AX 4	0.98	0.98	0.85	0.88	0.85	0.88	0.82	0.79							0.96	0.95					0.89
AX 5	0.99	0.99	0.99	0.99	0.92	0.96	0.98	0.96							0.74	0.77					0.93
AX 6																	0.84	0.89			0.87
CO 1	1	1																			1.00
CO 2			0.98	0.98	0.98	0.99	0.98	0.89	0.97	0.98	0.99	0.98									0.97
CO 3																			0.63	0.83	0.73
SA 1		1		0.99						1		0.99		1							1.00
SA 2	1		0.99						1		0.99		1								1.00
계																					0.93

0% 100%

산학 협력 프로젝트

- 제출 실적
  - SW 제작 (산학 1), 국제학술대회 논문 1편 이상 게재 (산학 2), 국내 특허출원 1건 이상 (산학 2)
- 현재 진행 사항
  - SW 등록: 진행중
  - 특허 출원: 문서 제작 거의 완료
  - 국제 논문 작업 시작할 계획

References

Barber, Rina Foygel, and Emmanuel J. Candès. "Controlling the false discovery rate via knockoffs." *Annals of Statistics* 43.5 (2015): 2055-2085.

Ghorbani, Amirata, Abubakar Abid, and James Zou. "Interpretation of neural networks is fragile." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 2019.

Lu, Yang Young, et al. "DeepPINK: reproducible feature selection in deep neural networks." *arXiv preprint arXiv:1809.01185* (2018).