

Attention を応用した鏡面反射光の非活性化による CNN 推論精度改善手法

松永 聖明^{1,a)} 赤嶺 有平^{†1} 根路銘 もえ子^{†2}

概要: 博物館など室内にて展示品を展示する施設では、照明による鏡面反射光が展示品や背景に生じやすい。特に表面が滑らかな展示品にはその傾向が顕著に見られ Convolutional Neural Network(CNN) をベースとするモデルで上記の物体を識別対象とした場合、学習時その物体特有の特徴領域を上手く捉えられず背景や鏡面反射光自体を特徴として捉え学習を進めてしまうことが見受けられる。これはモデルの誤識別を引き起こす原因の1つであり、解決されるべき問題である。そこで本研究では Attention 機構をベースとする CNN の推論根拠可視化手法に着目し、入力画像中より特定した鏡面反射光領域を元にモデルの特徴マップより該当ピクセルの非活性化を行うことで鏡面反射光による見えの変化を含む画像に対しても正常な学習を行う画像認識モデルを考案した。本稿では博物館の展示品を対象とした自作データセット及び CIFAR100 のベンチマークデータセットを用いて本手法の有効性に関する検討を行い、その検証結果より先行手法に比べ一定の精度の改善が認められた。

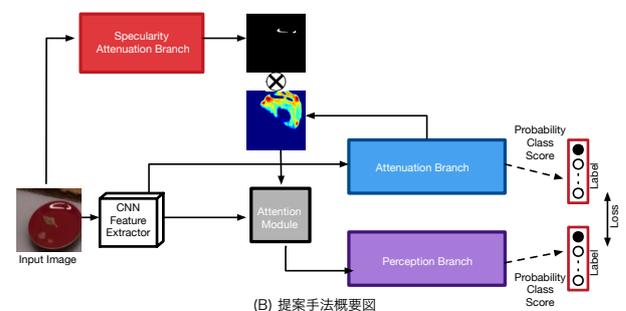
1. はじめに

2012年を皮切りに Convolutional Neural Network(CNN) をベースとするモデルは画像認識を始め、物体検出など様々な分野で高い精度を出している [1][2][4][3][6][7][8][9]。

一方で CNN には依然として周辺照明の変化による物体の見えの変化 (Appearance Changes) に影響を受け易いことが問題としてある。特に鏡面反射光は画像における物体の特徴を変えてしまうことから CNN の訓練時、背景や Specular Highlights 自体など対象物体固有の特徴ではない領域を特徴箇所として注視してしまう場合 (図 1)(A) がある。加えてデータセットが少量に限定されていた場合、これらの物体に対する CNN による識別はより困難なものとなる。そこで本研究では CNN における推論根拠の解釈可能性を明らかにする Explainable Artificial Intelligence(XAI)[18] の中でも Attention を用いて学習中モデルの注視領域を可視化可能な Attention Branch Network(ABN)[19] に着目し、



(A) 鏡面反射光を含む展示品(左)と学習済みCNNの注視領域例(右)
朱漆貝堆石鉢皿



(B) 提案手法概要図

図 1 提案手法概略図

モデル学習時、入力画像より特定した鏡面反射光領域を元に Attention 機構を用いて特徴マップにおける該当ピクセルの発火減衰を行うことで見えの変化を含む画像に対しても正常な学習を行う画像認識モデルを考案した (図 1)(B)。

本稿では始めに関連研究について紹介し、その後、本提案手法の詳細について述べ最後に実験と実験の結果について具体的に示す。本稿における実験では自作データセット及び CIFAR100 のベンチマークデータセットを用いて本手法の有効性についてを用いて検討を行い、実験結果より本手法を適用することで鏡面反射光を含む入力画像に対し先行手法に比べ一定の精度向上が認められた。

¹ 琉球大学大学院 理工学研究科
University of the Ryukyus
the Graduate School of Science and Engineering
^{†1} 琉球大学 工学部
Presently with University of the Ryukyus
Presently with the Department of Engineering
^{†2} 沖縄国際大学 経済学部
Presently with Okinawa International University
Presently with the Department of Economics
a) 145165b@ie.u-ryukyu.ac.jp

2. 関連研究

2.1 鏡面反射光領域の検出及び除去

通常人間が視覚から物を識別する際、脳は赤・緑・青のRGB色空間ではなく、明度・彩度・輝度のHSV色空間で各数値を推定している。そこでFranciscoら[25]は、輝度が高く、彩度が低いピクセルを鏡面反射光領域として定めRGBの入力画像より輝度・彩度を算出後、独自の計算式MS-diagramを利用して鏡面反射光特定を実現している。また彼らの手法では特定した鏡面反射光領域に該当するピクセルを入力画像より直接取り除くモジュールを用意している。

一方で本手法では同手法を取り入れ入力画像より鏡面反射光に該当するピクセルを特定するものの除去はせず、代わりにCNNにおける特徴マップにおいて該当する箇所が活性化しない処置を行うことで実質的に鏡面反射光領域を注視しない識別器を設計した。

2.2 CNN 推論根拠の可視化

現在画像認識タスクにおいてCNNをベースとするモデルを対象とし注視領域を可視化する説明可能なAIの研究には大きく

- 順伝播情報を用いる手法 (top-down 方式)
 - 逆伝播情報を用いる手法 (bottom-up 方式)
- が存在する。

top-down方式の可視化手法では学習済みモデル後段における全結合層を取り除き、代わりにGlobal Average Pooling(GAP)層を接続することで推論によるカテゴリ毎の注視領域をヒートマップとして取得するClass Activation Mapping[13]が提案されている。一方bottom-upの可視化手法では、CAMの可視化領域に加え、カテゴリ毎の勾配情報を畳み込み層の出力から獲得して利用し、推論根拠を可視化するGradient-weighted Class Activation Mapping(Grad-CAM)[14][15][16]及びGuided Backpropagation[10][11][12]などが存在する。しかしこれらの可視化手法は学習済みモデルへの適用を想定されており学習中においてモデルの注視領域を獲得することはできない。

そこで学習時にモデルの注視領域が獲得できる手法としてAttention機構[20]を利用したAttention Branch Network(ABN)[19]が提案されている。入力画像中における特徴抽出をVGG16等の学習済みモデルで行い、Attention Branch及びPerception Branchの2つのモジュールを用いて学習を進め、従来の画像分類モデルにおける推論根拠可視化手法と異なりネットワークの学習中においても注視領域を獲得することに成功している。

しかし同研究を問題解決に応用することを想定した場合、ネットワークの構造上、特徴量抽出器が鏡面反射領域を注視して学習を進めることで見えの変化による影響を増幅させ

てしまう可能性がある。従って本研究では[19]のAttention Mapに対し、[25]の手法によりSpecular Highlightsと判断された領域を非活性化する操作を加えることで同領域を学習しないように制御する。これによりモデルの識別精度改善が期待できる。

3. 提案手法

本節では提案手法について詳細に述べる。

図(2)に示すように本手法は以下4つのモジュールを統合したネットワーク構造となっている。

- Feature Extractor
- Attention Branch
- Specularity Attenuation Branch
- Perception Branch

本手法において入力画像を始めに受け取るFeature Extractorでは、VGG[2]・ResNet[3]・InceptionV3[4]などの学習済みモデルにおける全結合層を取り除き特徴抽出器として扱うことで画像に含まれる物体の特徴領域を特徴マップから出力として獲得する。

3.1 Attention Branch

Attention BranchではFeature Extractorにより獲得した物体の特徴箇所において、各クラスに寄与度の高い特徴箇所をAttention Mapとして取得する。これは図(2)に示したようにAttention Branchの最終層を全結合層ではなく、 $C \times 3 \times 3$ の畳み込み層に置き換えることで $C \times W \times H$ の特徴マップを取得していることが要点となる。ここで C はクラス数を、 W 及び H はカーネルを表している。取得した特徴マップをGlobal Average Poolingにより $C \times 1 \times 1$ の次元にダウンサンプリングし各クラス確率を計算することで空間的な特徴を失うことなく各クラスごとの発火領域を保持した特徴マップが生成され、これにより学習中においても視覚的にモデルの推論根拠を把握することが可能となる。

さらにAttention Branchでは中間層より分岐して得た特徴マップに対しSigmoid関数を適用することで物体以外の特徴領域を注視しないよう促すAttention Mapを生成しAttentionモジュールを通じてPerception Branchに渡る特徴マップに掛け合わせることで推論精度を向上させる役割を担う。

このヒートマップは図(2)の $J(x_i)$ に該当する。

しかしこのネットワーク構造では、前段の特徴量抽出器であるFeature Extractorが鏡面反射領域を注視して学習を進めていた場合、これを増幅させた特徴マップ出力をPerception Branchに届けてしまう。そこでSpecularity Attenuation Branchより鏡面反射光領域を注視しない学習を促すAttenuation Mapを生成し、Attention Mapに掛け合わせることで、モデルの鏡面反射光による見えの変化に

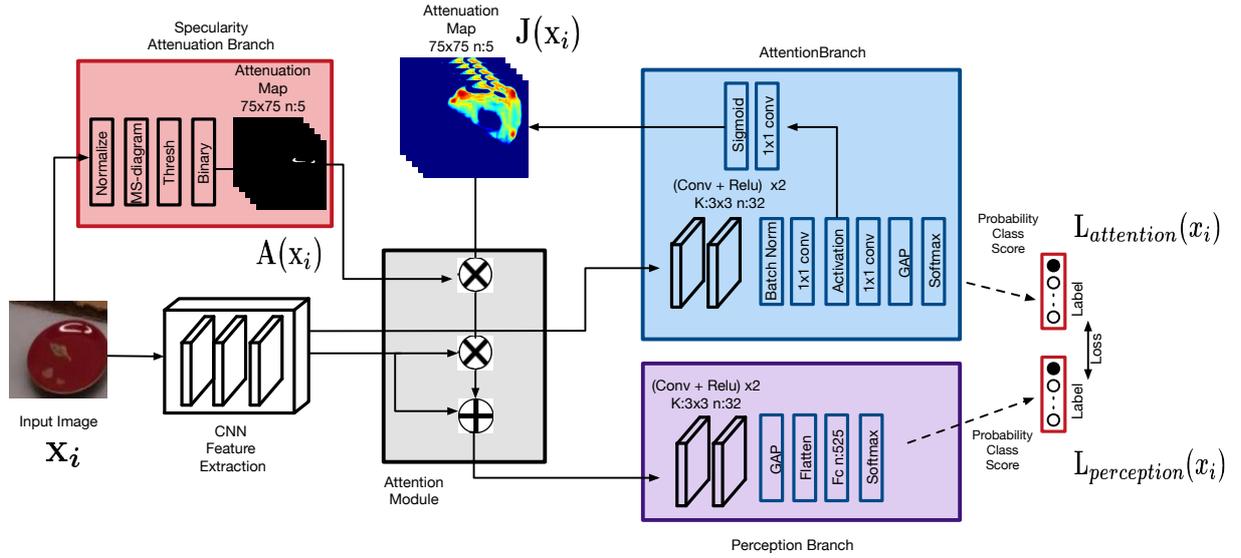


図 2 提案手法詳細図

対する影響を軽減する。

3.2 Specularity Attenuation Branch

Specularity Attenuation Branch では入力画像に含まれる鏡面反射光領域の特定を行い、特定した領域を元に鏡面反射光に反応している特徴マップピクセルを非活性化する Attenuation Map を生成する。この Attenuation Map を Attention Branch にて獲得した Attention Map に対し乗算することでネットワーク全体の特徴マップに対し鏡面反射光領域に反応しないよう発火減衰を行うことを実現する。始めに入力画像における鏡面反射光領域のの特定に関して本手法では F. Ortiz らの鏡面反射光特定手法 [25] を利用する。

具体的には式 (1) にあるように入力画像を正規化して明度を獲得し

$$m = \frac{1}{3} * (r + g + b) \quad (1)$$

式 (2) にあるように彩度を算出する。ここでの (r, g, b) は RGB 色空間の座標を表しておりそれぞれ $r \in [0, 255]$, $g \in [0, 255]$, $b \in [0, 255]$ である。

$$s = \begin{cases} \frac{1}{2}(2r - g - b) = \frac{2}{3}(r - m) & \text{if } (b + r) \geq 2g \\ \frac{1}{2}(r + g - 2b) = \frac{3}{2}(m - b) & \text{if } (b + r) < 2g \end{cases} \quad (2)$$

そこで式 (3) にあるように F. Ortiz らの提案する MS-diagram を用いて入力画像に含まれる鏡面反射光領域を特定する。

$$f(m, s) = \frac{\log(MS(m, s)) * 255}{\operatorname{argmax}(\log(MS(m, s)))} \quad (3)$$

その検出結果を図 (4)(下段) に示す。

続いて同手法にて獲得した鏡面反射光領域を特定した画像 $f(m, s)$ を 2 値化し鏡面反射に該当する領域及びそれ以

外の領域のピクセルを次の式 (4) を用いて変換する。

$$A_{uv} = \begin{cases} 0, & \text{if } (f_{uv}) == 255 \\ 1, & \text{elif } (f_{uv}) == 0 \end{cases} \quad (4)$$

この操作により本ネットワークの特徴マップにおける鏡面反射光を含む領域の非活性化可能な Attenuation Map を生成する。

3.3 Perception Branch

Perception Branch ではネットワークにおける最終的なクラス予測を行う役割を担う。Feature Extractor, Attention Branch, Specular Attenuation Branch よりそれぞれ対象物体を含む Feature Map の獲得, Attention Map, Attenuation Map を Attention モジュール内にて式 (5) のように統合し出力 $\hat{F}_c(x_i)$ を Perception Branch の入力とすることで最終的なクラス確率を算出する。

$$\hat{F}_c(x_i) = J_c(x_i) * A(x_i) * F_c(x_i) + F_c(x_i) \quad (5)$$

ここで $F(x_i)$ は Feature Extractor による特徴マップ出力を、 c はカテゴリを意味する。

結果として Perception Branch では鏡面反射光領域領域を無視し対象物体の特徴領域を注視する特徴マップを入力として学習が進められるためより正常な学習が期待できる。

3.4 ネットワークの学習

ネットワークの学習には式 (6) に示すように、入力画像 x_i に対し Attention Branch, Perception Branch の 2 つの損失関数における和を取る $\hat{L}(x_i)$ を用いて学習を進める。

$$\hat{L}(x_i) = L_{\text{attention}}(x_i) + L_{\text{perception}}(x_i) \quad (6)$$

式 (6) における $L_{\text{attention}}(x_i)$ は Attention Branch における推論に対する損失を、 $L_{\text{perception}}(x_i)$ は Perception

Branch における損失を各表しており、本稿における実験では損失関数はクロスエントロピー誤差を採用した。

4. 実験

4.1 実験方法

本実験では鏡面反射光を含みやすく見えの変化を生じやすい物体を対象とした画像分類タスクに対し本手法を適用することで

- 推論精度が改善されるか
- 物体固有の特徴を注視し正しい学習が行われるか

の2つの観点から博物館の展示品を対象とし自作したデータセット及び CIFAR100 のベンチマークデータセット [27] を用いて本手法の有効性に関する検証を行う。入力画像は 300x300 ピクセルにリサイズし正規化を行った後に本ネットワークの学習に利用する。この検証により入力画像における物体に現れる見えの変化による影響を抑える事で識別精度が向上するか、モデルの有効性について検討する。また本手法では Feature Extractor として自作のネットワーク BaseNet(表 1) を利用するが VGG16, ResNet50, InceptionV3 などの学習済みモデルを利用することでさらに精度を改善できることが見込まれる。故に実験では各の学習済みモデルを転移学習させ比較検証を行う。

Feature Extractor CNN (BaseNet)		
Kernel size & Layer name	Layer description	
Input 300 × 300 RGB image		
3x3 conv	32	ReLU
3x3 conv	32	ReLU
2x2 MaxPooling		
0.25 DropOut		
3x3 conv	5	ReLU
2x2 MaxPooling		

表 1 BaseNet

4.2 データセット

実験で使用するデータセットに関して、本実験では独自で撮影・収集した博物館の展示品を対象とするデータセット及び CIFAR100 データを用いて本手法の有効性に関する検証を行う。博物館で収集した画像データセットに関して、外観の酷似する展示品 plate A, plate B, plate C の3カテゴリで構成されており特に鏡面反射光が含まれており見えの変化を生じている皿の展示品を対象に画像を収集した(図(3))。画像総数は 18000 枚であり、これらを train, validation, test に 12000:3000:3000 枚の割合で分割しネットワークの学習に利用する。

CIFAR100 のデータセットからは鏡面反射光による見えの変化を含みやすい4カテゴリ (bottle, plate, cup, can) を独自に選出し新たなデータセットとする。本実験では絵画

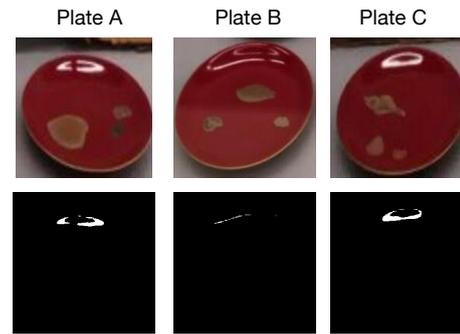


図 3 博物館における鏡面反射光を含む展示品 (上段) に対する鏡面反射光領域の特定 (下段)



図 4 CIFAR100 における鏡面反射光を含みやすい訓練画像 (上段) に対する鏡面反射光領域の特定 (下段)

像数 2400 枚を 1600:400:400 枚の割合で train, validation, test に分割しネットワークの学習に利用する。ネットワークの最適化について、本実験では勾配の大きさに応じて学習率を調整する RMSProp を用いてネットワークの最適化を行い、学習 epoch は 50, Batch Size は 8, 初期学習率は 10^{-3} に設定した。

本実験における全てのモデルは Tensorflow Keras で実装し、Nvidia GeForce GTX 1080 Ti を用いて学習を行った。

4.2.1 画像分類タスクにおける精度指標

本手法の精度指標として実験では

- Categorical Accuracy
- Recall
- Mean Square Error

の3つの指標を用いて本手法の評価を行う。また数値に対する信頼性を高めるためテストデータに対する評価は 10epoch 繰り返し各々の指標を元に得た数値の中央値を集計する。

5. 実験結果及び考察

本節では実験に対する実験結果を示し、

- 分類器としての精度改善評価
- 物体固有の特徴を注視し正しい学習の実行

についての2点を中心に本手法の有効性について検討する。また鏡面反射光の検出手法に関する改善についても検

討する。

5.1 推論精度の比較検証

鏡面反射光による見えの変化を含む画像の分類タスクに対する本手法の精度評価に関して実験では自作データセット及び CIFAR100[27] を用いて鏡面反射光を含みやすい物体を対象に本手法の有効性を検証した。

実験結果を以下 (表 (2)) に示す。

Dataset: CIFAR100			
Models	accuracy	mse rate	recall
VGG16 [2]	0.78	0.0646	0.172
InceptionV3 [4]	0.89	0.036	0.188
BaseNet+ABN [19]	0.675	0.1	0.085
VGG16+ABN [19]	0.86	0.044	0.193
InceptionV3+ABN [19]	0.89	0.041	0.2
BaseNet+Ours	0.712	0.07	0.167
VGG16+Ours	0.8175	0.057	0.194
InceptionV3+Ours	0.905	0.035	0.2

表 2 識別精度比較

実験結果として表に得られたように本手法は先行研究に比べ一定の改善が成していることが確認された。この結果から、本手法により鏡面反射光領域を非活性化することは全体の識別精度を向上することに対しある程度有効であると考えられる。また想定転していた通り学習済みモデルの転移学習により本提案手法によるモデルの推論精度をより高めることができると認められた。

5.2 推論可視化による評価

以下図 (5) 及び図 (6) にそれぞれ ABN[19] が鏡面反射光を含む入力画像に対し推論を行った場合及び本提案手法が推論を行った場合のモデルを用意し各の注視領域を Grad-CAM を用いて可視化した結果を示す。左列より入力画像, Grad-CAM, Guided-Back-Propagation の可視化を表している。

図にあるように ABN による推論結果では鏡面反射光領域及びプレート表面の模様箇所の双方に強く反応しており正しい学習が行われていないことが認められる。一方で本提案手法による推論結果 6 では鏡面反射領域を注視することなく、展示品のプレートにおける模様箇所 (特徴領域) を注視して学習を進めていることが確認でき、見えの変化による影響を受けない推論を行うことができるモデルが実現できている。

5.3 鏡面反射光領域特定手法の検討

乗算による特徴マップの非活性化では Attention Map おいて鏡面反射光に該当するピクセルの発火を 0 に置き換える。この処理では強制的に活性が抑制されるため、重要な

情報も同時に削除されてしまう可能性が拭きれない。故に更なる頑健性を持つ手法の実現にはこの処理の再考が必要となると考えられる。

5.4 Attenuation Map 生成に関する検討

本実験では Feature Extractor として学習済みモデルによる転移学習を用い検証を行った。デフォルトの BaseNet による検証では Attenuation Map サイズを (75x75x5) で生成したが転移学習 VGG16 のモデルではネットワークが深層なことから特徴マップ出力が畳み込まれることで縮小し (9x9x5) を, InceptionV3 のモデルでは (8x8x5) の Attenuation Map を生成し, Attention Map に乗算することで特徴マップ出力の非活性化を行っている。

ここで実験結果を参照すると BaseNet+ABN 及び BaseNet+Ours の精度改善の差が転移学習を行った場合に比べ大きくなっていることが確認できる。ここには Attenuation Map が縮小することにより、より広範囲の領域を非活性化することが関係している可能性があると考えられ Attenuation Map のサイズに関して更なる追加検証が必要である。

また本手法では入力画像より鏡面反射光領域を特定した Attenuation Map を 5 チャンネル複製することで Attention Map とチャンネル数を合致させ乗算可能にしている。故に全ての Attention Map に対し最適な非活性化処理が行われていない可能性もあると考えられる。この点に関しても追加の検証が必要である。

6. おわりに

本稿では、Attention Module による CNN の推論根拠ヒートマップを応用し入力画像における Specular Highlights 領域の発火減衰を実現する手法を考案し、その有効性について議論した。

結論として Specular Highlights による見えの変化を含む画像を識別対象とする場合、本手法を利用することで一定の精度改善が見込まれる可能性があることが確認された。加えて根拠可視化ヒートマップにおいても、物体特有の特徴箇所に発火している Attention Map の取得が確認され、正常な学習が実現できていると考えられる。

一方で、鏡面反射領域の減衰方法としては他にも様々な手法の適用が考えられ、更なる改善手法が必要であると考えられる。

謝辞 本研究は JSPS 科研費 19K01142 の助成によるものである。

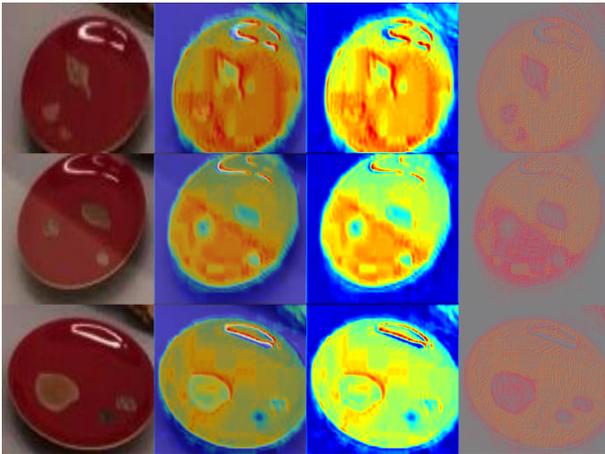


図 5 Grad-CAM[14] による学習済み ABN の推論根拠可視化

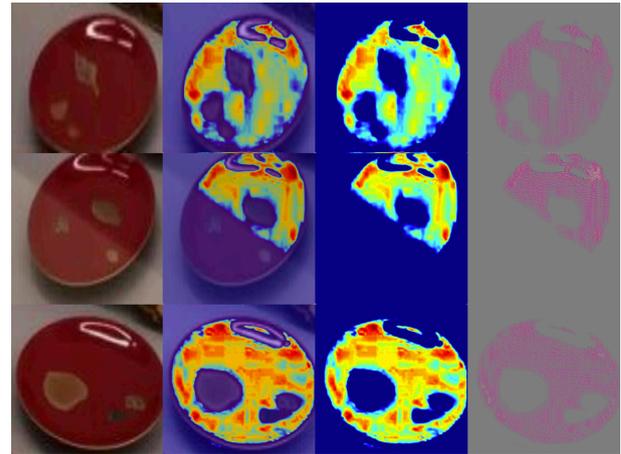


図 6 Grad-CAM[14] による学習済み提案手法の推論根拠可視化

参考文献

- [1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton : ImageNet Classification with Deep Convolutional Neural Networks, In Neural Information Processing Systems, 25(2), (2012).
- [2] Simonyan, Karen and Zisserman, Andrew: Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, (2014)
- [3] K. He, X. Zhang, S. Ren, and J. Sun: Deep residual learning for image recognition, Computer Vision and Pattern Recognition, 770–778, (2016).
- [4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna: Rethinking the Inception Architecture for Computer Vision, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2818-2826, (2016).
- [5] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik: R-CNN Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, IEEE Conference on Computer Vision and Pattern Recognition, 580-587, (2014).
- [6] R. Girshick:Fast R-CNN, 2015 IEEE International Conference on Computer Vision (ICCV),1440-1448,(2015).
- [7] Ren, S. He, K., Girshick, R. B., Sun, J.: Faster R-CNN towards Real-time Object Detection with Region Proposal Networks, In NIPS 2015, 91-99., (2015).
- [8] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg : SSD Single Shot MultiBox Detector, Computer Vision ECCV 2016 pp 21-37, (2016).
- [9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi:YOLO (You Only Look Once) Unified, Real-Time Object Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 10.1109/CVPR.2016.91, (2016).
- [10] Zeiler, M. D., Fergus, R: Visualizing and Understanding Convolutional Networks, European Conference on Computer Vision, 818–833, (2014).
- [11] Karen Simonyan, Andrea Vedaldi, Andrew Zisserman: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, Workshop at International Conference on Learning Representations, (2014).
- [12] J.T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller: Striving for Simplicity: The All Convolutional Net, ICLR (workshop track), (2015).
- [13] Zhou, Bolei and Khosla, Aditya and Lapedriza, Agata and Oliva, Aude and Torralba, Antonio: Learning Deep Features for Discriminative Localization, Computer Vision and Pattern Recognition, (2016).
- [14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, IEEE International Conference on Computer Vision (ICCV), pp. 618-626, (2017).
- [15] A. Chattopadhyay, A. Sarkar, P. Howlader and V. N. Balasubramanian: Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks, 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 839-847, (2018).
- [16] Wang, Haofan and Wang, Zifan and Du, Mengnan and Yang, Fan and Zhang, Zijian and Ding, Sirui and Mardziel, Piotr and Hu, Xia: Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks, The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, (2020)
- [17] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje: Learning important features through propagating activation differences, In ICML, 3145–3153, (2017).
- [18] Erico Tjoa and Cuntai Guan: A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI, IEEE transactions on neural networks and learning systems, (2020).
- [19] Hiroshi Fukui, Tsubasa Hirakawa, Takayoshi Yamashita, Hironobu Fujiyoshi: Attention Branch Network: Learning of Attention Mechanism for Visual Explanation, Computer Vision and Pattern Recognition, 10705-10714, (2019)
- [20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin: Attention is all you need, arXiv preprint arXiv:1706.03762, (2017).
- [21] J. Hu, L. Shen and G. Sun: Squeeze-and-Excitation Networks, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,7132-7141,(2018).
- [22] Liangzhi Li and Bowen Wang and Manisha Verma and Yuta Nakashima and Ryo Kawasaki and Hajime Nagahara: SCOUTER: Slot Attention-based Classifier for Explainable Image Recognition, arXiv, (2020).
- [23] Residual Attention Network for Image Classification: Wang, Fei and Jiang, Mengqing and Qian, Chen and

- Yang, Shuo and Li, Cheng and Zhang, Honggang and Wang, Xiaogang and Tang, Xiaoou, arXiv preprint arXiv:1704.06904, (2017)
- [24] P. Wang and N. Vasconcelos: SCOUT: Self-Aware Discriminant Counterfactual Explanations, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8978-8987, (2020)
- [25] F. Ortiz and F. Torres: Automatic detection and elimination of specular reflectance in color images by means of MS diagram and vector connected filters, In IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 36, no. 5, pp. 681-687, (2006).
- [26] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei: ImageNet: A large-scale hierarchical image database, In IEEE CVPR, 248-255, (2009).
- [27] Alex Krizhevsky: Learning multiple layers of features from tiny images, (2009).