DiDA: 機械学習による画像診断の判断ポイントを 入力差分により解析・可視化する手法

綾塚 祐二^{1,2,a)} 雅樂 隆基^{1,b)} 安川 力^{3,c)} 吉高 淳夫^{2,d)}

概要:機械学習を用いた画像分類は医療を含め幅広い分野で多くの成果を上げている。しかしその分類は 何を根拠としてなされているかが人間には判りづらい場合も多く、それを可視化するためのさまざまな研 究が行われている。疾患の分類のような目的のためには、画像中の「それらしい」正の寄与部分(所見)だ けでなく「そぐわない」、すなわち負の寄与部分となる所見も診る必要があるが、既存の研究では負に寄与 する部分は可視化の対象として注目されていない。我々は、機械学習モデルに対し、画像に微小な差異を 加えた画像を入力した場合の確信度の出力の大きさの変化を画像化し提示する手法、DiDA を提案する。 提案手法ではグリッド単位で区切りマスクした画像を用いて出力の差異を捉え、複数のグリッドサイズを 用いることで、正負の寄与領域を的確に描出する。本稿では、DiDA を光干渉断層計による眼底の断層画 像からの疾患分類に適用した例を挙げる。眼科医の見解と照合した結果、提案手法による解析画像は正負 の寄与を的確に捉えていることが判った。

1. はじめに

畳み込みニューラルネットワーク (Convolutional Neural Network, CNN)を始めとする機械学習を用いた画像分類は 幅広い分野で多くの成果を上げている。その一つが医療の 分野であり、著者らのグループでも光干渉断層計 (Optical Coherence Tomography, OCT) による眼底の断層画像から 眼底疾患の種類を分類する研究 [1] などを行っており、こ うした分類が精度良く行えることを確認している。

一方で機械学習による分類は、何を根拠として分類され たかが人間には判りづらい場合も多く、それを改善するた めに「説明可能な AI」(Explainable Artifitial Intelligence) と呼ばれる取り組みが行われている。画像分類においても アテンションマップと呼ばれるものなど、画像中の注目し た領域を可視化する方法の研究が既に数多くなされてきて いる。しかし、既存の研究ではある分類に属するオブジェ クトや領域を適切に抽出することを主眼としており、眼底 の断層画像からの眼底疾患の分類のような、症状や状態な どの所見の組み合わせにより総合的に判断すべきような事 例には不十分である。特に「この疾患だとするとここに見 える異常所見は合わない」というような負の寄与が扱えないという問題がある。

「疾患を正確に診断するために、症状が他の要因で起 こっていることを否定し、可能性のある疾患を絞り込む」 ことは鑑別診断と呼ばれるが、画像分類においても鑑別の 補助となる情報を提供することで医師の診断をサポートで きると考えられる。そのためには機械学習による画像の各 分類に対する、画像中の正の寄与部分・負の寄与部分をそ れぞれ明確に視覚化する必要がある。

我々はその具体的な手段として、ある画像を分類した機 械学習モデルに対し、その画像に微小な差異を加えた画像 を入力した場合の出力値 (分類に対する確信度)の変化の 方向と大きさを画像化し提示する手法、DiDA (Differential Image Diagnostic Analysis)を提案する。具体的には、画 像をグリッドに区切り、一ヶ所ずつグリッドをマスクした 画像を機械学習モデルに分類させ出力の変化の方向(正負) で色分けしたものをそのグリッドの色とし、出力の変化の 大きさを輝度に反映した画像を生成する。複数のサイズの グリッドに対してそのような画像を生成し、重ね合わせる ことで画像中のさまざまな大きさの特徴を捉え可視化する ことができる。

2. 関連研究

画像分類を行う機械学習モデルの分類根拠の可視化を行 う既存の研究のアプローチは大別して

¹ 株式会社クレスコ 技術研究所

² 北陸先端科学技術大学院大学先端科学技術研究科

³ 名古屋市立大学大学院 医学研究科 視覚科学

^{a)} ayatsuka@acm.org

 $^{^{\}rm b)}$ t-uta@cresco.co.jp

 $^{^{\}rm c)} \ \ {\rm yasukawa} @med.nagoya-cu.ac.jp$

 $^{^{\}rm d)} \quad ayoshi@jaist.ac.jp$



図1 元の入力画像とグリッドに区切り一ヶ所ずつマスクした画像

(1) CNN の中間層のデータから画像を生成する

(2)入力に撹乱を与えた場合の出力値から画像を生成する(3)分類モデルとは別の説明生成モデルを構築する

に分けられる。以下に述べるように、どのアプローチでも 分類ターゲットはオブジェクトあるいは領域として表され るものであることが (暗黙の) 前提となっており、ターゲッ トがどこに存在するかを可視化しようとしている。すなわ ち、たとえば画像中のどこに「犬」がいるのかを可視化し ようとしており、それがなぜ (「猫」ではなく)「犬」と判 断されたのかという情報を提示するものではない。

2.1 CNN の中間層のデータから画像を生成する

DeepLIFT[2] は CNN の分類結果を出力するセルから入 力に向かって backpropagate することによって元の画像 上へ結果を反映するような形で画像を生成する。CAM[3] や Grad-CAM[4]、Grad-CAM++[5] では畳み込み層の最 終段の出力のエッジの重みや、activation map の変化に よる出力の変化から画像 (attention map) を生成してい る。Score-CAM[6] は activation map のチャネルを考慮し attention map の精度の改善を試みている。

これらは CNN というアーキテクチャに依存した手法で あり、また正の寄与部分のみが扱われている。Attention mining と呼ばれる、attention map の改善を試み、分類の 精度自体も向上させる研究([7], [8], [9] など)でも分類対 象の領域を取りこぼしなく抽出することが主眼となってお り、負の寄与に関しては扱われていない。

Bach らの研究 [10] では、負の寄与部分が可視化されて いる例が挙げられている。しかし、手書き文字の分類を対 象としており、分類対象の文字として想定される形状と合 致しない部分が示されているにすぎない。

2.2 入力に撹乱を与えた場合の出力値から画像を生成する

Petsiuk らの RISE[11] は、画像をグリッドに区切りラン ダムに数ヶ所マスクしたものを多数用意し、それぞれのマ スク画像を学習済みの機械学習モデルに対し入力した際の (対象となる分類の) 出力を重みとしてそれらのマスクを 足し合わせたものを説明画像とする手法である。入力をグ



図 2 元画像を分類したときとマスクした画像を分類したときの出 力値の差分を色と輝度にマップする

リッド単位でマスクするのは我々のアプローチと同じであ るが、複数箇所を同時にマスクした画像に対する出力値を 重みとしており、結果として正の寄与部分しか取り出すこ とができず、可視化されるのは対象の存在する領域となっ ている。

グリッド単位で正の寄与部分を集めるのはいわば積分的 なアプローチと言えるのに対し、我々の手法はグリッド単 位での正ならびに負の変化を調べる微分的なアプローチで あると表現することもできる。RISE では複数のグリッド サイズを用い特徴の大きさを捉えることも想定されておら ず、その点も我々の手法との差異と言える。

2.3 分類モデルとは別の説明生成モデルを構築する

LIME[12] や SHAP[13]、Anchors[14] などは、分類する 対象を「特徴のセット」として表し、各特徴の有無が分類 にどう影響するかを調べることにより説明モデルを構築す る。これらの研究では画像も扱われているが、画像をどの ように「特徴のセット」として表すかには触れられておら ず、予め何らかのセグメンテーションが行われていること を前提としている。我々の手法では、セグメンテーション が困難な、境界の曖昧な特徴なども探索可能である。

Residual Attention Network[15] や Attention Branch Network[16] のように、CNN による分類モデルから分 岐させたネットワークにより分類と同時に attention map を生成させるというアプローチもある。これも、正の寄与 部分だけを扱っているという点は他のアプローチと変わら ず、負の寄与に関しては考慮されていない。

3. DiDA のアルゴリズム

我々の提案する DiDA は、分類器への入力と、それに 対する出力値のみから解析画像を生成する。そのため、出 力値がそれぞれの分類に対する確信度やその分類らしさ を表すものであれば、機械学習によるものを含め分類の 仕組みを問わず適用が可能である。以下、具体的な解析画 像の生成手順を説明する。まず、グリッドサイズを設定し て DiDA Grid Image を生成し、複数のグリッドサイズの DiDA Grid Image から、DiDA Mixed Image を生成する。



図 3 同じ入力画像、同じ分類モデルに対して生成された、グリッド サイズだけが違う 3 つの DiDA Grid Image (赤いグリッド は正の寄与部分、青いグリッドは負の寄与部分である)

3.1 DiDA Grid Image

図 1 左上の画像を、分類及び解析を行いたい画像 I_o と する。これに対して、同図右側の画像群のように、任意の サイズ (図の例では 5×5 個に分けるサイズ) のグリッドに 区切りーヶ所ずつマスク (図では黒一色の塗りつぶし) し た画像を用意する。(i,j) の位置のグリッドをマスクした 画像を M_{ij} とする。画像 I をある分類器 C で分類したと きのカテゴリ k に対する確信度の出力の値を $C_k(I)$ と表 す。出力値は個々の分類に対する出力そのままの値でも、 全分類の出力に対し Softmax など何らかの正規化を行っ た後の値でもよい。

このとき、画像 I_o の (i, j) の位置のグリッドをマスク することによるカテゴリ k に対する出力値の変化 d_{ij} は次 のように表される。

$$d_{ij} = C_k(I_o) - C_k(M_{ij})$$

この値は、マスクした箇所にカテゴリ k に対し正に寄与す る特徴が存在するならば正の値となり (すなわち、マスク した画像を入力した際の確信度は下がる)、負に寄与する特 徴が存在するならば負の値となる (マスクした画像を入力 した際の確信度は上がる) はずである。

生成する画像を G とし、これを入力画像に対するグリッ ドと同じ数のグリッドに区切る。(i,j)の位置のグリッド は d_{ij} から、 $|d_1| > |d_2|$ ならば $a(d_1) > a(d_2) \ge 0$ となる ような性質を持つ関数 a により $a(d_{ij})$ で表される輝度 (最 大値を超える場合は最大値)の単色で塗りつぶす。色は例 えば、 $d_{ij} \ge 0$ ならば赤、 $d_{ij} < 0$ ならば青のように設定す る (図 2)。このようにして全てのグリッドに彩色したもの が DiDA Grid Image (以下、Grid Image と略す)である。

図3に挙げたのはある同じ入力画像、同じ分類モデルに 対して生成された、グリッドサイズだけが違う3つのGrid Image である。正の寄与部分、負の寄与部分がはっきりと 分けられ可視化されているのが判る。細かいグリッドサイ ズのほうが特徴の位置や形状を解像度よく捉えることがで きることが見てとれる。

しかし、粗い 7×7 のグリッドで正の寄与度が高い明るい 表示色になっている画像右側、中央やや下の領域が、細か



図 4 8 つのグリッドサイズの DiDA Grid Image から生成された DiDA Mixed Image

い 21×21 のグリッドでは全て暗い色となっている。これ は、細かなグリッドでは当該部分の正の寄与度を持つ画像 中の特徴が一部しか隠されず、分類の確信度があまり下が らないためと解釈できる。すなわち、複数のサイズの Grid Image を較べることで、画像中の特徴のスケールを推定す ることができると考えられる。

3.2 DiDA Mixed Image

前述のように、Grid Image は提示できる情報がグリッ ドサイズに応じて変わる。そこで複数の Grid Image を統 合し、一枚の画像に情報をまとめて提示することを考える。 その画像を、DiDA Mixed Image と呼ぶ (以下では Mixed Image と略す)。

まとめ方にはさまざまなバリエーションが考えられる が、ここでは Grid Image を重み付けしつつ重ね合わせる 方法を採用する。すなわち、Grid Image *G* の座標 (x, y)のピクセル値を p(G, x, y) と表すと *n* 枚の Grid Image G_i (i = 1, 2, 3, ..., n) から重み w_i で生成される Mixed Image D の座標 (x, y) のピクセル値 p(D, x, y) は

$$p(D, x, y) = \sum_{i=0}^{n} w_i \cdot p(G_i, x, y)$$

となる。

細かなグリッドのほうがやや重めになるようにして、7×7 ~ 21×21 の 8 つの奇数サイズの Grid Image から生成さ れた Mixed Image を図 4 に示す。重みの付け方は一例で あり、対象や抽出すべき情報に合わせて調整することを想 定している。

4. DiDA の適用例: 眼底断層画像の分類

我々は DiDA を用いて、OCT で得られた眼底の断層画 像を分類する機械学習モデルが妥当な分類を行っている かどうかを検証するためのアプリケーション DiDA View を試作した。対象とした機械学習モデルは、ニデック社の OCT 機器、RS-3000 Advance 及び RS-330 で撮影された 画像を、32 種類の診断名と健常、合わせて 33 種に分類す



図 5 DiDA View による表示例 (1)

るものであり、垂直・水平断面併せて 3,173 枚の教師画像 を、横幅を 1/2 に縮め*1 224×224 ピクセルにフィットす るようにスケーリングしたものを EfficientNet(B0)[17] に 学習させたものである。学習済みのモデルは、与えられた 画像 (学習データと同様、224×224 ピクセルに調整された もの) に対する各分類の確信度のスコアをそれぞれ出力す る。793 枚のテスト画像を分類させた場合の、正答率 (確 信度が最も高くなるものが正答である率) は 72.1%、確信 度上位 3 つ以内に正答が入る率は 88.9%である。

図5はDiDA View の画面例である。4つの画像が表示 されており、最左は元の画像及び熟練した医師による診断 名(この場合はDry AMD (萎縮型加齢黄斑変性))である。 その右に並ぶのは、元の画像の上に Mixed Image を半透明 で重ね合わせた画像であり、機械学習モデルが出力した確 信度の高い上位3つの分類(この場合は順に、Dry AMD, RA (網膜萎縮), normal (健常))に対応するものである。そ れぞれ右下部にスコアが表示されている。中央の小さな正 方形は、画像中の同じ部分を比較しやすいよう、ユーザの 操作したマウスカーソルの位置に合わせて表示される。

それぞれの分類に対し、赤 (正の寄与部分) と青 (負の寄 与部分) が、共通する部分も違う部分もあるパターンで分 布していることが判る。図 5 の画像中の正方形で囲われ た部分は、Dry AMD の分類には負に寄与しており、「Dry AMD らしくない」と機械学習モデルが判断していると解 釈できる。RA に対してはその周辺を含め強めの赤、正の 寄与の表示が出ており「RA の特徴的所見のように見える」 ことを示している。normal に対しては暗めの青色となっ ている。「『健常』に対する負の寄与部分」とはすなわち「何 らかの病変」と解釈することができる。

図6に他の4つの例を示す。図5の例も含め、機械学習 の判断が複数の診断名に対して拮抗したスコアを出してい る症例を抽出した。眼底疾患に詳しくない者にとっても、 注目すべきであろう箇所が明確に判るので、医師のための 診断補助だけでなく、患者への説明用途や、初学者に対す る教育用途として応用することも考えられる。

5. 眼科医の見解との照合

前節で提示したような DiDA Mixed Image は、熟練し た眼科の医師の読影によりラベル付けしたデータにより学 習した機械学習モデルの出力から構築されている。学習に より、同じラベルがつけられた画像に共通して存在する特 徴が見いだされ、それに基づき分類を行っているはずであ り、DiDA はそれを可視化しているはずである。しかし、 医学的に見て妥当な情報が DiDA により示されているかど うかは慎重に検証する必要がある。

その端緒としてまず、図 5,6 で挙げた例に関して、学 習データのラベル付けを行った眼科の医師である筆者らの 一人が検証した。その結果 DiDA で提示している情報は 医師が判断する際のポイントとおおよそ合致していること が確認できた。ここではそのうち 3 つの症例について説明 する。

5.1 萎縮型加齢黄斑変性 (Dry AMD) の症例

図 5 の症例に対し医師は Dry AMD と判断し、機械学 習は Dry AMD と RA のスコアが拮抗している。この画 像 (最左) について医師が Dry AMD と判断したポイント は、画像中央付近の下部の明るくなっている部分である。 これは地図状萎縮と呼ばれる部位の網膜色素上皮 (正方形 部分の中央やや下を左右に横切る明るい線状の部分に相 当、図 7 参照) の萎縮により OCT の観察光がその下の脈 絡膜 (図 7 参照) まで透過しやすくなっており、脈絡膜の 輝度が高く写っているということである。Dry AMD の診 断においては、この輝度が高くなる境界部位と、萎縮部位 以外の網膜は正常に近い所見であるところを見ている。一 方、RA の場合 (右から二枚目)、まさに網膜の萎縮による 菲薄化を見ていると考えられる。

Dry AMD の画像では脈絡膜の明るくなっている部分の 左側がやや青くなっている。この部分に関して医師は「意 識していなかったが、Dry AMD であれば脈絡膜がもっと 菲薄化している症例が多く、この症例の脈絡膜厚は例外的」 と述べている。DiDA により無意識には見ているであろう 部分を顕在化させている例と考えられる。

^{*1} 眼底の OCT 画像は元々縦 (奥行き) 方向が強調されたスケール になっており、この比率でも医師は違和感なく読影が行える



図 6 DiDA View による表示例 (2)



図7 網膜色素上皮と脈絡膜

5.2 黄斑浮腫 (ME) の症例

図 6(a) は医師が ME と診断名をつけた症例である。ME は疾患名ではなく網膜の肥厚を示す異常所見の名称であ り、浮腫の部分に内部が液体 (黒色) の嚢胞様所見を伴うと CME (嚢胞様黄斑浮腫) と呼ばれる。医師の判断は、ME は認められるが嚢胞は顕著ではなく、疾患の特定もできな い、というものである。機械学習では ME と CME で判 断が割れ、次点で BRVO (網膜静脈分枝閉塞症) という疾 患名を出力している。

正方形で囲われている、CME の画像 (右から二枚目) で 局所的に強く赤色で示されている黒い空洞が医師は明確で ないと判断した嚢胞様腔である。ME の画像 (左から二枚 目) ではその周辺も含め赤色が示されている。これは、小 さな嚢胞様所見よりも網膜全体の肥厚 (浮腫) を見て ME と判断していると考えられる。これは「対象となるオブ ジェクトや領域を抽出する」のとは違う可視化が必要であ ることを示す例であると言える。

最右の BRVO は、症状として CME を伴うことも多い。 画像右側中央付近に薄く赤で示されている、輝度の高い小 さないくつかの点は硬性白斑またはその前駆体を示唆する 高輝度病変 (hyper-reflective foci) と呼ばれるものであり、 BRVO の症状として現れ得るものである。また、BRVO は 上下に別れて分布する網膜静脈枝の閉塞であるため、上下 (画像の左右) で浮腫の偏りがあることが特徴である。本画 像は水平断であるが、画面左半分の網膜が正常であること を確認していることが赤色で示されている (図 6(b) の症例 の右から二枚目の画像も同様)。これらの所見を総合して 機械学習は BRVO を挙げていることが判る。

5.3 加齢黄斑変性前駆病変 (Early AMD) の症例

図 6(d) は加齢黄斑変性 (AMD) の前駆病変と分類され る症例である。機械学習による Early AMD の判断の画像 (左から二枚目) で顕著な赤い表示の部分 (正方形で囲われ た部分とその周囲) は網膜色素上皮の隆起が孤発性に存在 し (凹凸周囲は異常がない)、ドルーゼンと呼ばれる加齢性 沈着物であるとして前駆病変と診断される所見である。

一方、Wet AMD (滲出型加齢黄斑変性)の特徴である脈 絡膜新生血管も同様の網膜色素上皮の隆起を認めるため、 右から二枚目の画像において赤い表示となっているが、そ の左右には青い領域が見られる。Wet AMD であれば隆起 がより広範囲に渡るか、あるいは滲出性網膜剥離という症 状を伴うことが多く、この青い領域はそれらが見られない ことを示していると考えられる。最右の normal の判断は、 異常所見に乏しい画像であることを示しているが、その中 でも、網膜色素上皮の隆起は青色すなわち正常らしくない (異常)所見と判断されている。

6. まとめ

本稿では、ある画像を分類するモデルに対し、その画像 に微小な差異を加えた画像を入力した場合の出力の値の変 化を画像化し提示する手法、DiDA を提案し、眼底断層像 からの機械学習による眼底疾患の分類に応用した例を紹介 した。正の寄与部分と負の寄与部分の双方を明確に区別し て提示できることが大きな特徴であり、また、入力する画 像と出力の関係だけを扱うため、分類モデルの種類によら ず適用することができる汎用性の高い手法である。眼底断 層像の分類に対し DiDA を適用した結果が、医師が判断す る際のポイントを的確に示していることも確認した。

今後、CAM など既存の関連研究で提案されている手法 で可視化した場合との差異や得失も含め、適切な可視化が 行えているかどうかの検証を行っていく。また、DiDA に より生成された画像が医学的に適切な情報を提示できてい るかをより詳細に検証し、有効性を確認する。加えて、こ うした情報を医師や患者などに提示し活用する手法につい ても検討を行う。

機械学習を始めとする AI による診断などは、「中身が ブラックボックスだ」として不安がられることも多い。し かし、本提案のような手法を通して判断のポイントなどを 適切に可視化することができれば、機械学習はむしろ、暗 黙知を含む人間の持つ知識を外在化させ、つぶさに解析す るための手段として活用することができると考えられる。 我々は、そのような可能性も追求してゆく。

参考文献

 Kuwayama, S., Ayatsuka, Y., Yanagisono, D., Uta, T., Usui, H., Kato, A., Takase, N., Ogura, Y. and Yasukawa, T.: Automated Detection of Macular Diseases by Optical Coherence Tomography and Artificial Intelligence Machine Learning of Optical Coherence Tomography Images, *Journal of Ophthalmology*, Vol. 2019, DOI: doi:10.1155/2019/6319581 (2019).

- [2] Shrikumar, A., Greenside, P. and Kundaje, A.: Learning Important Features Through Propagating Activation Differences, arXiv:1704.02685 (2019).
- [3] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. and Torralba, A.: Learning Deep Features for Discriminative Localization, arXiv:1512.04150 (2015).
- [4] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, *International Journal of Computer Vision*, Vol. 128, No. 2, p. 336359, DOI: 10.1007/s11263-019-01228-7 (2019).
- [5] Chattopadhay, A., Sarkar, A., Howlader, P. and Balasubramanian, V. N.: Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks, WACV 2018, DOI: 10.1109/wacv.2018.00097 (2018).
- [6] Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P. and Hu, X.: Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks, arXiv:1910.01279 (2020).
- [7] Zhang, X., Wei, Y., Feng, J., Yang, Y. and Huang, T.: Adversarial Complementary Learning for Weakly Supervised Object Localization, arXiv:1804.06962 (2018).
- [8] Wei, Y., Feng, J., Liang, X., Cheng, M.-M., Zhao, Y. and Yan, S.: Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach, arXiv:1703.08448 (2018).
- [9] Singh, K. K. and Lee, Y. J.: Hide-and-Seek: Forcing a Network to be Meticulous for Weakly-supervised Object and Action Localization, arXiv:1704.04232 (2017).
- [10] Bach, S., Binder, A., Montavon, G., Klauschen, F., Mller, K.-R. and Samek, W.: On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation, *PLOS ONE*, Vol. 10, No. 7, pp. 1–46, DOI: 10.1371/journal.pone.0130140 (2015).
- [11] Petsiuk, V., Das, A. and Saenko, K.: RISE: Randomized Input Sampling for Explanation of Black-box Models, arXiv:1806.07421 (2018).
- [12] Ribeiro, M. T., Singh, S. and Guestrin, C.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier, arXiv:1602.04938 (2016).
- [13] Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, *NIPS'17*, USA, Curran Associates Inc., pp. 4768–4777 (2017).
- [14] Ribeiro, M. T., Singh, S. and Guestrin, C.: Anchors: High-Precision Model-Agnostic Explanations, *Proceed-ings of the AAAI-18, the IAAI-18, and the EAAI-18,* AAAI Press, pp. 1527–1535 (2018).
- [15] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X. and Tang, X.: Residual Attention Network for Image Classification, arXiv:1704.06904 (2017).
- [16] Fukui, H., Hirakawa, T., Yamashita, T. and Fujiyoshi, H.: Attention Branch Network: Learning of Attention Mechanism for Visual Explanation, CVPR '19 (2019).
- [17] Tan, M. and Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, *Proceed*ings of the 36th International Conference on Machine Learning, Vol. 97, PMLR, pp. 6105–6114 (2019).