人間科学研究　第32巻第２号　217－223（2019）

原 著 論 文

Convolutional Neural Networkを用いた区画数検出の検討

古賀 敬之，高橋 友太，岡崎 俊太郎，大須 理英子

Detecting the Number of Compartments Using Convolutional Neural Network

Takayuki Koga, Yuta Takahashi, Shuntaro Okazaki and Rieko Osu

(Faculty of Human Sciences, Waseda University)

(Received：May 13，2019 ; Accepted：August 1，2019)

Abstract

In this study, we examined the compartment counting ability of Convolutional Neural Network（CNN）using images that are controlled for their features（i.e. brightness, shape, and the number of lines that divided compartments）. We trained the CNN by images with limited features and tested its generalizability by images that have features different from those of training data in the shape combination and shape itself. Consequently, the CNN achieved approximately 56% accuracy in the generalization test, which was higher than chance level of 33%. This result indicated that the CNN learned to count the division of compartments without identifying the shape of the compartments. However, the CNN was not able to count the division of compartments included the shapes which were qualitatively different from that of training data. The acquired counting ability in CNN was limited compared to that of animals.

Key Words：Neural Network, Concept Acquisition, Counting, Number Sens

１．はじめに

人間はナンバーセンス（数感覚）を持っており数の大きさを視覚情報から直感的に捉えることができるため物の形状や配置が変わってもおおよそ数を数えることができる[1][2]．例えばリンゴが３つであってもミカンが３つであっても同じ３という数を認識することができる．また合わせて６つと捉えることも可能である．

一方で，最近では機械学習によって画像の中にある物体の数を数える（カウンティング）の研究が報告されている．例えば，Convolutional Neural Network（CNN）を用いて画像や動画から人の群衆の数や車の台数の推定を行った研究などが報告されている[3][4][5]．またCNNの構造や入力データに対して工夫をすることでその精度を向上させることができると報告されている[6][7]．しかしこれらは

早稲田大学人間科学学術院（*Faculty of Human Sciences, Waseda University*）

人の顔や車など特定の物体に対して数を推定している．しかし作成されたCNNでそれ以外の物体が数えられるかどうかを検証していないので，このようなCNNが生物の数感覚のような，汎化性のあるカウンティング能力を獲得したかどうかは明らかではない．

そこで本研究では，CNNのカウンティング能力に汎化性を持たせられるかどうかを検討した．CNNによるカウント機能の汎化性を確認するためには，異なる形状の対象に対してその数のカウントを学習し，さらに，学習時に与えられていない形状を含む対象に対しても数の推定が可能であるかどうかを検証する必要がある．その際以下の2点に留意した．１点目はCNNの特性を考慮した画像を用いたことである．CNNの畳み込み層1層目はエッジ検出の機能に関わっており，複雑なエッジを正確に検出するためには畳み込み層に十分なチャンネル数，チャンネルサイズが必要である [8]．本研究の目的であるCNNのカウント機能の検証には，エッジ検出の難易度は関係しないため，エッジ検出が容易な画像を用いて実験を行った．具体的には，3本以内の直線(全て水平方向または垂直方向)によって様々なパターンで２～４区画に分けられた画像を作成し，それらの画像からCNNが異なる形状の区画が含まれていてもその数をカウントすることができるようになるかどうかを検討した．２点目は使用画像の特徴パターンを可能なかぎり統制したことである．本実験での検証のためには学習時やテスト時の画像データは教師ラベルである区画数以外の特徴（例えば直線数や輝度）には大きな違いがない統制された画像であることが望まれた．そこで，直線の太さをランダムに設定し，区画数の違いにより画像全体の輝度にばらつきが出ないように統制した．また，外枠を除く直線による交点の数が，いずれの区画数においても0個から２個になるようなパターンで区画が分割された画像を作成した（図１）．これらの統制により，画像輝度や直線数，交点数の違いによって区画数が判別できないようにした．

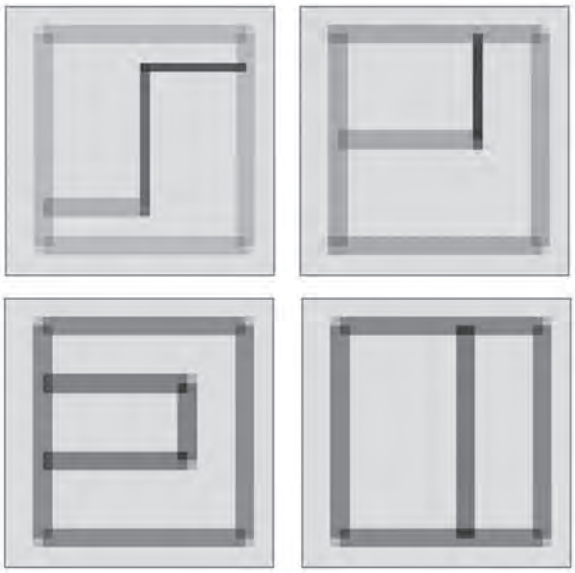
モデルの汎化性を確認するため，学習時には与えられなかった形状の区画が存在する画像を新たに作成し，学習後のモデルで区画数の推定を行いその精度を確認する汎化テストを行った． CNNが特定の特徴ではなく，汎化性のあるカウンティング能力を獲得した場合，学習していない形状の区画に対しても区画数カウントを行うことができると考えられる．そこで，汎化テスト用の画像に対して33%（本実験でのチャンスレベル），あるいはより保守的な基準である50%（詳細は後述）を超えた精度で推定ができれば，CNNが汎化性のあるカウンティング能力を獲得できているものとした．カウンティング汎化性は学習時の画像パターンに影響を受けるため [9]，汎化テストで用いる画像の区画形状と学習時の区画形状との類似度によって推定精度にばらつきが生じることが予想される．そこで汎化テストには，学習に用いる画像との類似度に考慮した複数パターンの画像を用いた（図３，４）．学習時の画像との類似度が高いほど推定精度は高くなることが予想される．数カウント概念以外に獲得された特徴による影響をなるべく排除するため，画像内の交点や直線の本数，学習時の画像からの可変性，輝度など汎化テスト用の画像においてもできるかぎり統制した．汎化テストの結果から，CNNがどのような特徴を抽出し区画数を推定しているのかを考察した．

２．実験方法

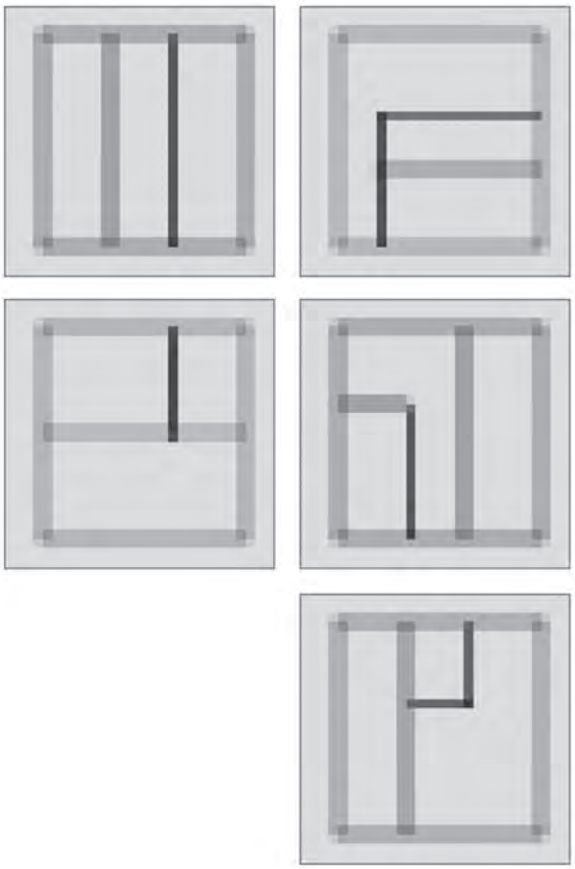
2.1. 学習方法

入力データは1チャンネル，28×28ピクセルの画像を使用した．３本以内の直線によって区画数が２～４個になるように作成し，それぞれ30000枚ずつ合計90000枚用意した．そのうち67500枚を訓練データに使用した．３本以内の直線で構成したものを学習に使用したため，それぞれの画像は複数のパターンで構成された．直線３本以内で構成するという統制のため全ての区画数において長方形（I字型）かL字型が含まれたが，区画数２のみU字型の区画が生じた（図１）．また，訓練データに使用していない22500枚を通常のテストデータとして使用した．

区画数２（I字型，L字型，U字型が含まれる）



区画数３（I字型，L字型が含まれる）



区画数４（I字型，L字型が含まれる）

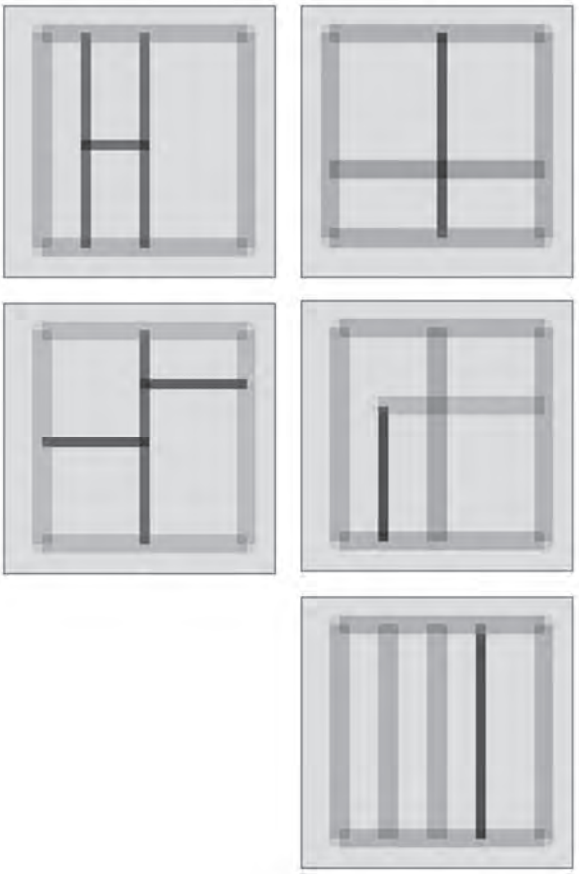


図１　訓練データに用いた画像パターン

本研究で使用したCNNは畳み込み層が２層，プーリング層が２層，全結合層が２層であった（図２）．畳み込み層でのカーネルサイズは５×５で，ストライドが１，１層目は16チャンネル，２層目は32チャネルであった．活性化関数には学習を高速化するため$f(x)=\max(0,x)$で定義されるRectified Linear Units（ReLU）を利用した[10]．最適化関数にはAdaptive Moment Estimation（Adam）を使用した．学習時の学習率は0.0001であった．また，本研究で使用したCNNは深層学習用フレームワークのPytorchを用いて作成した．学習用画像67500枚に対してバッチサイズは400，学習エポック数は20であった．この学習サイクルでモデルを10個作成した．10個のモデルを作る際，訓練データとテストデータは全て異なる分け方をした．

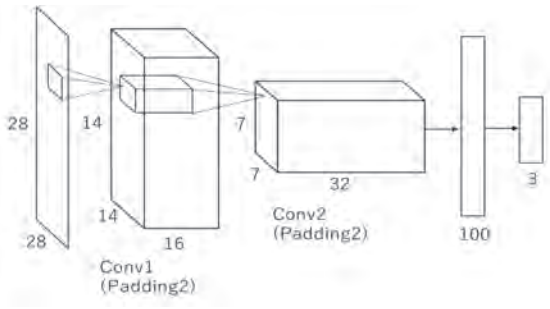
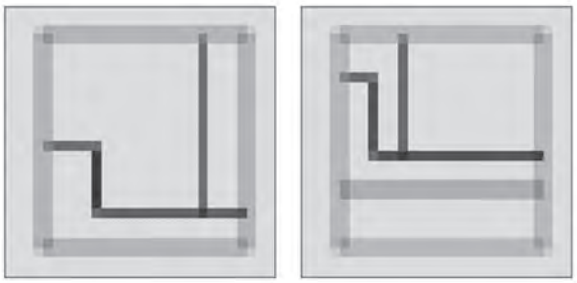


図２　CNNの構造

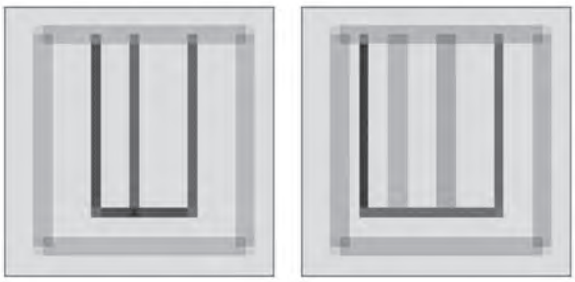
2.2. 汎化性のテスト方法

学習したモデルの汎化性を確かめるため学習時と同じ区画数でかつ直線４本以上使用する画像を新たに作成し，その画像を用いて精度を確かめた．その際，学習時に区画の形を特徴として獲得しているのかを確認するため，「学習時に与えられた既知の形の区画を使った新しいパターンの画像」と，「学習時には与えられなかった未知の形の区間を含むパターンの画像」の両方を作成し，それらの精度を比較した．区画の形が未知か既知かどうか以外の要因をなるべく排除するために，１）学習時に与えられたパターンから直線を足すことで成立する，２）直線が４本と５本である，３）区画数が増えると交点が２点増える，４）直線の幅をランダムに設定し輝度のばらつきがないようにする，という４点を統制して汎化テスト用画像を作成した（図３）．ただし，直線４本以上でかつ既知の区画で構成される区画数２の図形は作成できないため，汎化テスト用画像の区画数はすべて３と４で構成された．そのため汎化テストにおける精度は２択のチャンスレベルである50%をより保守的な基準とした．さらに，外枠と接しない区画を含むようなより新規性が高いと予想される画像を２パターン作成した（図４）．したがって，汎化テスト用の画像は５つのカテゴリに分かれた（表１）．

既知の区画で構成される新たなパターン１（LI型）



既知の区画で構成される新たなパターン２（U_1型）



未知の区画が含まれるパターン1（S型）

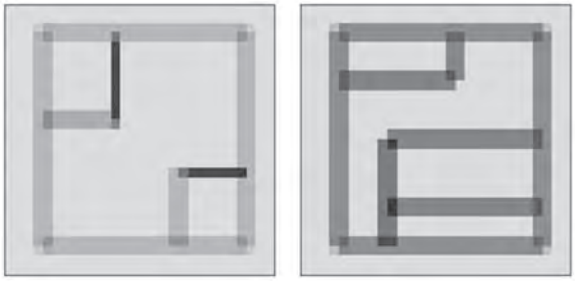
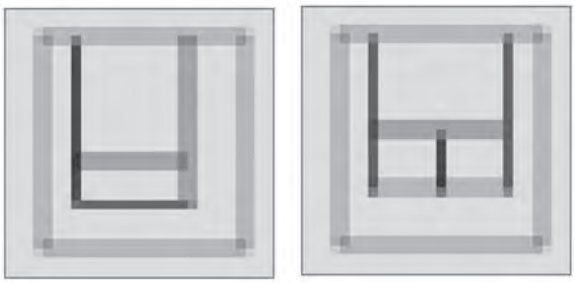


図３　汎化テスト用の画像パターン

既知の区画が含まれるパターン２（U_2型）



未知の区画が含まれるパターン２（O型）

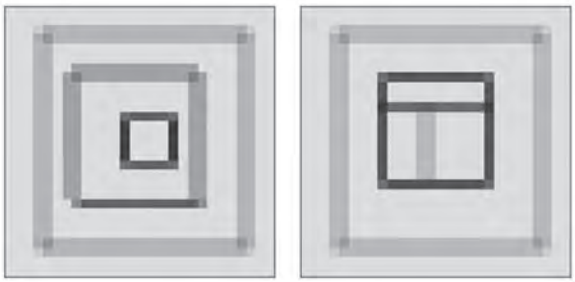


図４　外枠と接しない区画をもつ汎化テスト用の画像パターン

表１　各汎化テスト用画像のカテゴリ

カテゴリー	パターン	U字型	外枠と接しない区画
LI型	既知	なし	なし
U_1型	既知	あり	なし
S型	未知	なし	なし
U_2型	既知	あり	あり
O型	未知	なし	あり

３．結果

通常のテストデータに対しては全てのモデルが99%を超える精度になるように学習することができた（図５）．汎化テストでは，汎化テスト画像のカテゴリごとに精度の違いがみられた．LI型が最も精度が高く平均精度（再現率）は60.2%であった．次にS型，U_1型，U_2型という順で，精度はそれぞれ55.8%, 53.2%, 49.9%であった．最も精度が低かったのはO型で，36.2%であった（図６）．保守的な基準である50%以上の精度との比較では，LI型，S型，U_1型において有意に高かった（$p < .001$）．U_2型では50%と有意な差はなく（uncorrected $p = .697$）O型では33%よりも有意に高かった（$p = .013$）ものの50%と比べると有意に低かった（$p < .001$）．

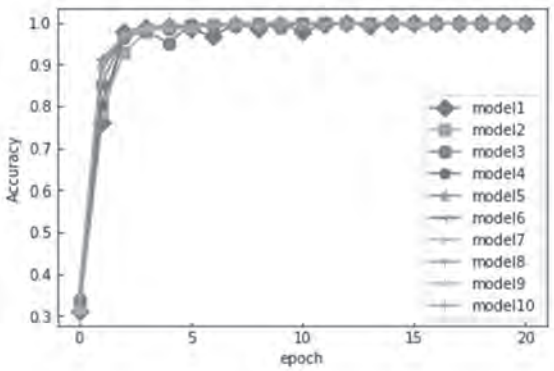


図５　通常のテストデータに対するエポックごとの推定精度

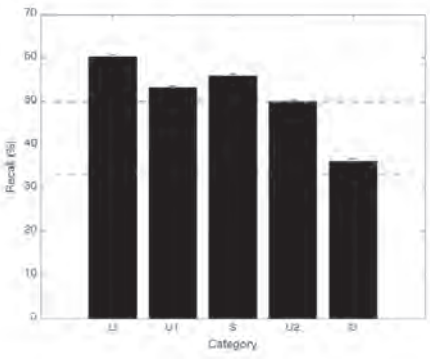


図６　汎化テストにおける各カテゴリの再現率

４．考察

本実験では，学習データと共通の特徴を持つ通常のテストデータに対しては，高い精度で区画数を推定できるモデルを作成できた（図５）．また汎化テスト用画像のパターンの違いによって精度が異なることが示された（図６）．LI型，未知の区画が含まれるS型，U_1型に対して推定精度が50%を有意に上回る結果であったことから，CNNが一定以上の汎化性をもつ区画カウント機能を獲得したことが示された．特に既知の区画で構成されるLI型に対して最も精度が高くなった結果に関しては，CNNが学習過程においてデータ駆動で特徴を抽出する性質[9]が関係していると考えられる．つまり訓練データの全ての区画数にL字型とI字型が含まれていたため，それらの形状を認識できるように学習できていたためだと考えられる．

一方で，外枠と接しない区画を含むU_2型とO型に対しての精度は50%以下であったことから区画が外枠と接しているかどうかが区画数カウントにおいて重要な特徴であったことが考えられる．さらにどの直線も外枠と接していないO型に対しての精度はより低く，３択のチャンスレベルである33%に近かったことから，直線と外枠との接点の数が本実験での学習データから区画数を推定すべき対象として検知する重要な特徴として抽出され，区画数の推定に使用されていたと考えられる．

本実験に加えて，ニューラルネットの全結合層を２層から５層に増やしたモデルや，同様の２層構造にドロップアウトを追加したモデル，２層目の畳み込み層のチャンネル数を32から64に増やしたモデルで同様の実験を行なったが結果に有意な違いは見られなかった．また，エポック数を減らして学習させたモデルでの実験においても結果に有意な違いは見られなかった．これらの結果から，本実験における汎化性能の向上において，CNNの層数やチャンネル数，学習回数は大きな影響は与えないことが推測される．

今回の研究では，U字型が区画数２にのみ含まれていたことによる影響が交絡要因となっているため，今後は画像カテゴリを十分に統制した上で，より詳細な検討が必要であろう．

生物がもつような数感覚は対象の形状に関わらず数の多さを直感的に認識できる能力である．しかし

本研究でのCNNはU_2型，O型に対しての区画数推定精度が低く，抽出した特徴が明らかでないため，数感覚が獲得されているとは言いきれない．一方で，より汎化性のある抽象的な概念を獲得することができれば，今後ニューラルネットワークによって生物がもつような数感覚を表現できる可能性があると考えられる．

また，数感覚と関連して数の概念がある．数の概念はPiagetが提唱した数に関わるより高次で人間的な概念である[11]．本研究でのCNNが数の概念を獲得した上で区画数カウントを行っていることは数感覚と同様に明らかではない．その理由は区画数カウント課題を解くにあたり，計算処理の過程が明らかでないためである．数の概念の獲得が達成されているとするならば，物体認識の問題に加えて対象の物体がどこにあるのかという物体検知の問題も解かれていなければならない．物体検知では画像から候補領域を抽出することと特徴の計算の処理が必要である[12]．しかし本研究で作成したCNNでは物体検知の要素に関してどこまで対応できているかについては不明である．したがって，数の概念の獲得可能性を議論するならば，今後はCNNが何を特徴として抽出しているのかを可視化し，詳細に検討することが必要である．

5．まとめ

本研究では，画像サイズや輝度，および区画形状や境界線を構成する直線の本数などの特徴を統制した画像データを用いて，CNNにおける区画数カウント機能について検討した．その結果，３択で区画数を推定することに関しては，学習時の全ての画像に共通する特徴（本実験では区画と直線が外枠と接すること）を持っている画像に対して未知の形状の区画が含まれていても50%以上の精度で推定をすることができた．つまり学習時の総データに共通する特徴を保持している画像に対してはCNNが一定以上の汎化性をもつカウンティング能力を獲得できたことが示された．

一方で，外枠との接点をもたない区画が存在する画像に対しては推定精度が50%に満たなかったことから，学習時のデータと大きく異なる特徴をもつ画像は区画推定の対象として認識されていない可能性が示唆された．また，本実験での学習データで学習したCNNは区画形状に全く依存せずに任意の形状の区画数を十分な精度でカウントできるまでには至っていなかったと言える．より高い汎化性能を達成するには本実験とは異なる学習データとモデルによる検討を要する．ただし，追加実験の結果から，汎化性能はモデル構造以上にデータ側に大きく影響を受ける可能性があるため，多くのパターンを含む画像データや，ノイズを加えた画像で学習を行うなど，学習データに対する検討がより重要であると考えられる．新しいモデルを検討する場合は層数や学習数ではなく，強化学習との組み合わせや空間情報を強力に保持できるとされるカプセルニューラルネットワーク[13]など根本的に新しい構造のモデルの検討が必要である．

文献

[1] S. Deheane, The Number Sense. 2015.

[2] V. Izard, C. Sann, E. S. Spelke, and A. Streri, “Newborn infants perceive abstract numbers,” *Proc. Natl. Acad*. Sci., vol. 106, no. 25, pp. 10382–10385, 2009.

[3] 池田浩雄, 大網亮磨, and 宮野博義, “CNNを用いた群衆パッチ学習に基づく人数推定の高精度化,” 情報科学技術フォーラム講演論文集, vol. 13, no. 3, pp. 105–106, 2014.

[4] C. Zhang, H. Li, X. Wang, and X. Yang, “Cross-scene crowd counting via deep convolutional neural networks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*., vol. 07-12-June, pp. 833–841, 2015.

[5] D. Oñoro-Rubio and R. J. L ▪ pez-Sastre, “Towards perspective-free object counting with deep learning,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*）, 2016.

[6] E. Walach and L. Wolf, “Learning to count with CNN boosting,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics*）, vol. 9906 LNCS, pp. 660–676, 2016.

[7] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y.

Ma, “Single-Image Crowd Counting via Multi-Column Convolutional Neural Network,” in 2016 IEEE *Conference on Computer Vision and Pattern Recognition* (*CVPR*) , 2016.

[8] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* , 2014.

[9] A. Mahendran and A. Vedaldi, “Understanding deep image representations by inverting them,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015.

[10] V. Nair and G. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *Proceedings of the 27th International Conference on Machine Learning*, 2010.

[11] J. PIAGET and M. M. Lewis, “LA NAISSANCE DE L’ INTELLIGENCE CHEZ L’ INFANT,” *Br. J. Educ. Psychol.*, 1939.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014.

[13] S. Sabour and G. E. Hinton, “Dynamic Routing Between Capsules,” no. Nips, 2017.