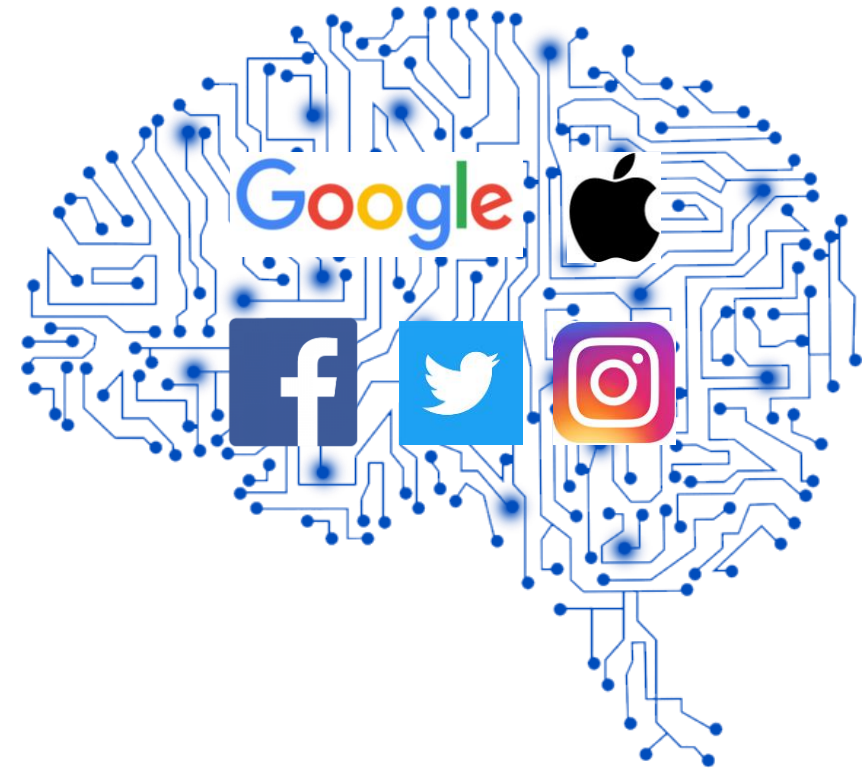


NHẬN DIỆN THƯƠNG HIỆU

LOGO DETECTION

Nhóm sinh viên thực hiện:

Nguyễn Ngọc Đàm	– KSTN Toán Tin K61
Nguyễn Phong Long	– KSTN Toán Tin K61
Ngô Thị Trà	– KSTN Toán Tin K61
Phạm Thị Thu Phương	– KSTN Toán Tin K61
Hà Thị Hào	– Toán Tin 01 K61



NỘI DUNG CHÍNH

- TỔNG QUAN
- XÂY DỰNG MÔ HÌNH
- TỐI ƯU
- KẾT QUẢ
- XÂY DỰNG ỨNG DỤNG

TỔNG QUAN

- **Tại sao cần nhận diện thương hiệu ?**
 - Cung cấp thông tin về việc công chúng tương tác với một thương hiệu cụ thể
 - Phục vụ phân tích truyền thông xã hội
- **Tại sao cần nhận diện thương hiệu qua ảnh ?**
 - Mạng xã hội phát triển
 - Ước tính 80% bài đăng trong phương tiện xã hội có chứa hình ảnh hoặc video
→ **Việc trích xuất thông tin có ý nghĩa từ văn bản và siêu tham số thông thường là không đủ !**
 - Hệ thống phần cứng đáp ứng đủ yêu cầu

TỔNG QUAN

Percentage of beer type in positive samples

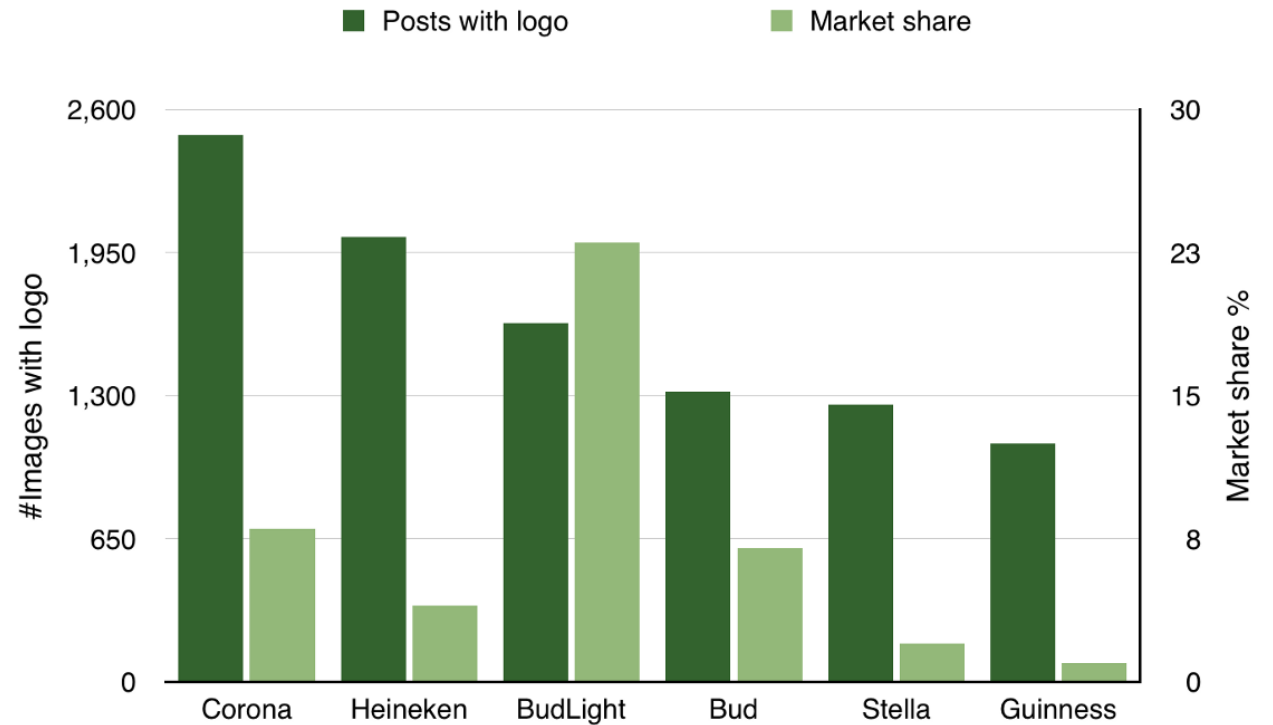
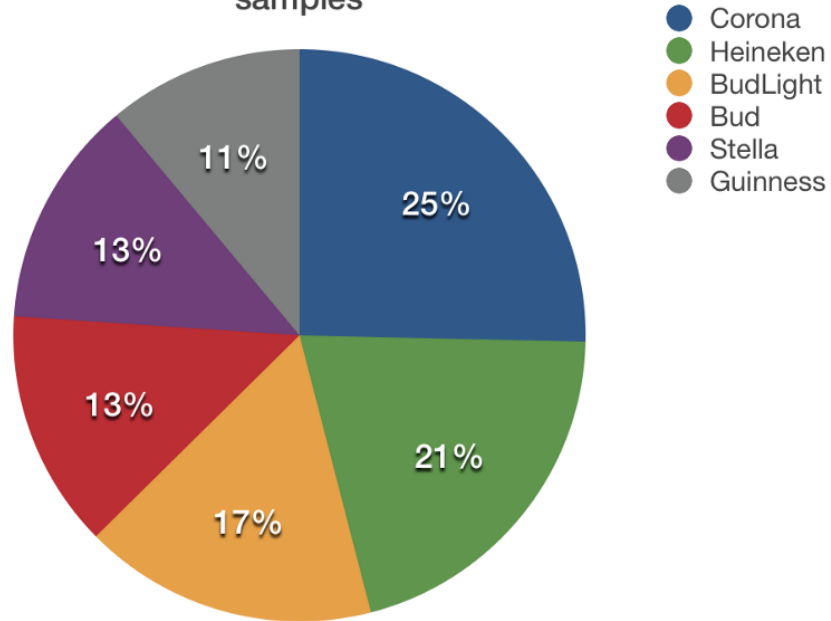


Fig 1: Phân tích truyền thông xã hội qua nhận diện logo từ ảnh

Nguồn: statics.com

TỔNG QUAN

- Vấn đề

- Các phương pháp xử lý ảnh thông thường kết hợp với máy học không đủ độ chính xác
 - Đặc trưng Haar-like và giải thuật Cascade
- Cần xây dựng mô hình mạng neural học sâu đảm bảo độ chính xác và thời gian tính toán
- Phương pháp sử dụng mô hình mạng neural học sâu cần nhiều chi phí tính toán, dữ liệu
 - 1-Stage: YOLO, SSD, ...
 - 2-Stages: Faster-RCNN, Mask-RCNN, Panoptic-FPN, ...



XÂY DỰNG MÔ HÌNH

- Kiến trúc 2-stages
 - Giai đoạn 1: Đề xuất vùng
 - Giai đoạn 2: Phân loại và hồi quy (Fast-RCNN)

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 1: ĐỀ XUẤT VÙNG

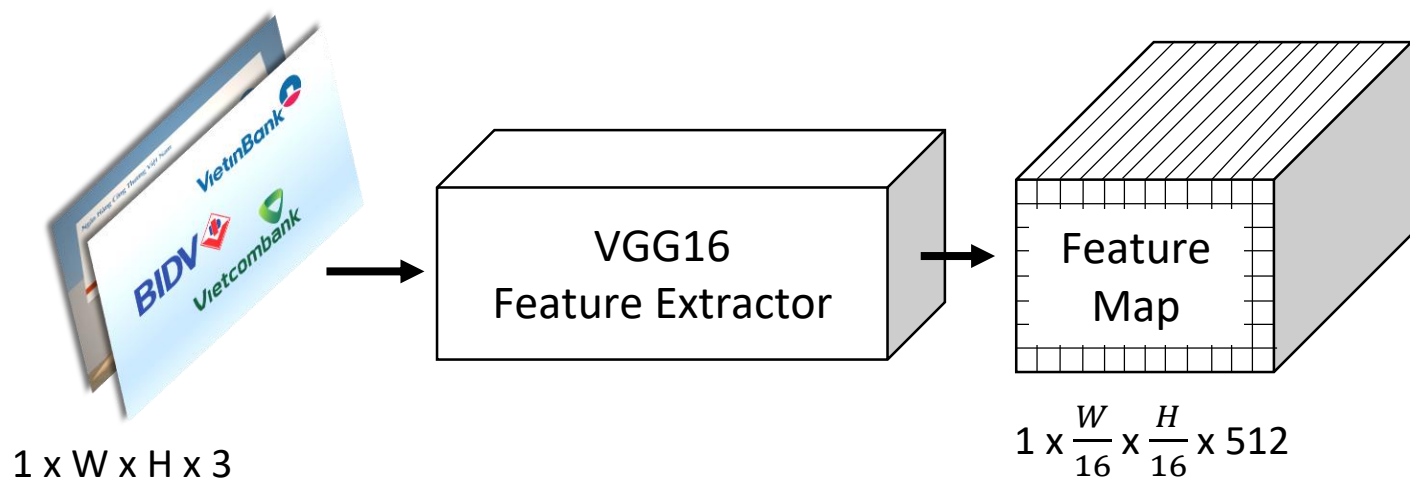


Fig 2.1: Trích xuất đặc trưng

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 1: ĐỀ XUẤT VÙNG

Kích thước features map giảm 16 lần so với ảnh gốc

→ Mỗi pixel trên features map mang thông tin của 16x16 pixels trên ảnh gốc



Fig 2.2: Trích xuất đặc trưng

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 1: ĐỀ XUẤT VÙNG

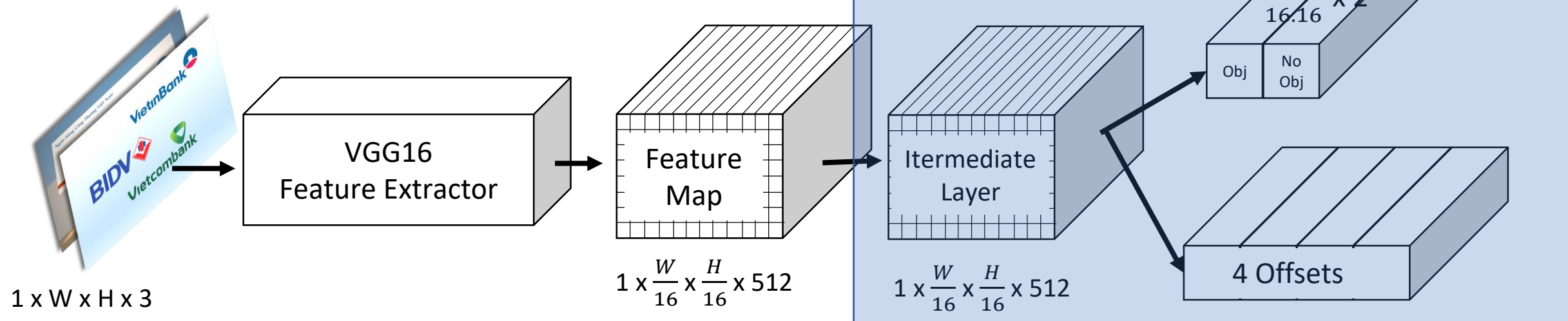


Fig 3: Mô hình đề xuất vùng

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 1: ĐỀ XUẤT VÙNG



Fig 4: Tập hợp anchors

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 1: ĐỀ XUẤT VÙNG

Anchors phản hồi với groundtruth là:

- Anchors có chỉ số IoU > 0.7
- Anchors có chỉ số IoU cao nhất

Anchors không phản hồi với bất kì groundtruth nào là:

- Anchors có chỉ số IoU với tất cả groundtruth < 0.3

Thực hiện lấy mẫu 64 anchors có phản hồi và 64 anchors không phản hồi



Fig 5: Tính phản hồi của anchors

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 1: ĐỀ XUẤT VÙNG



Input



RPN Output

Fig 6: Đầu vào, đầu ra của mạng đề xuất vùng

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 2: PHÂN LOẠI VÀ HỒI QUY

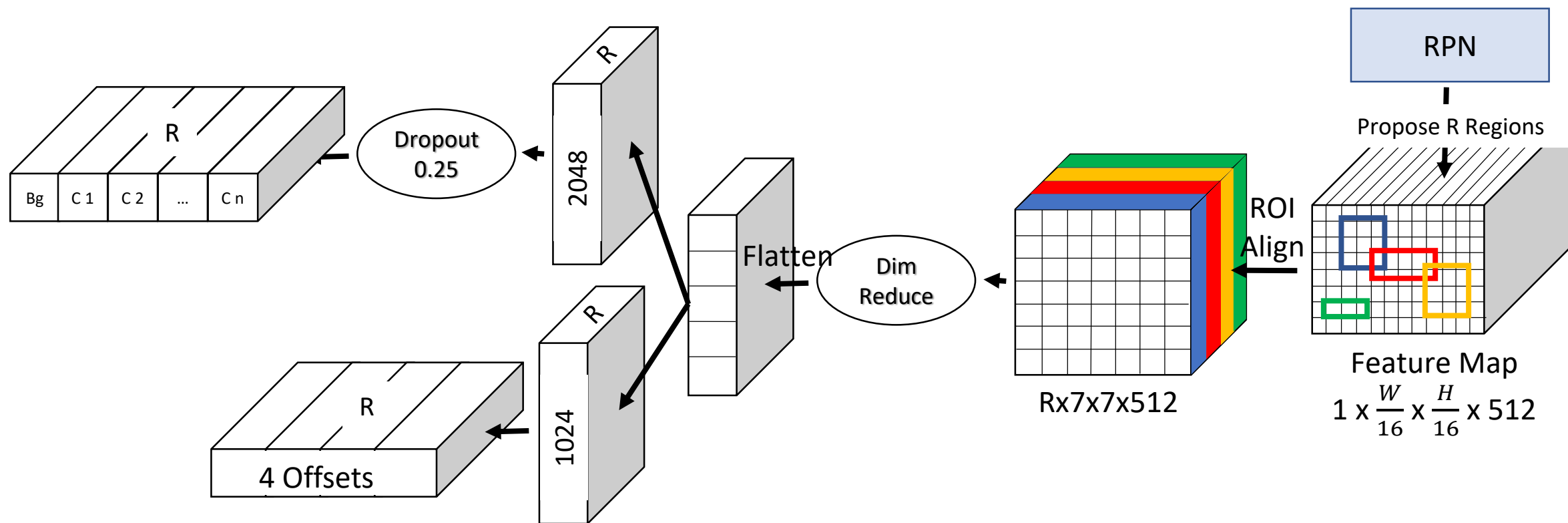


Fig 7: Mô hình giai đoạn 2

XÂY DỰNG MÔ HÌNH

- GIAI ĐOẠN 2: PHÂN LOẠI VÀ HỒI QUY

Chia sẻ tham số: Cả 2 giai đoạn sử dụng cùng 1 features map

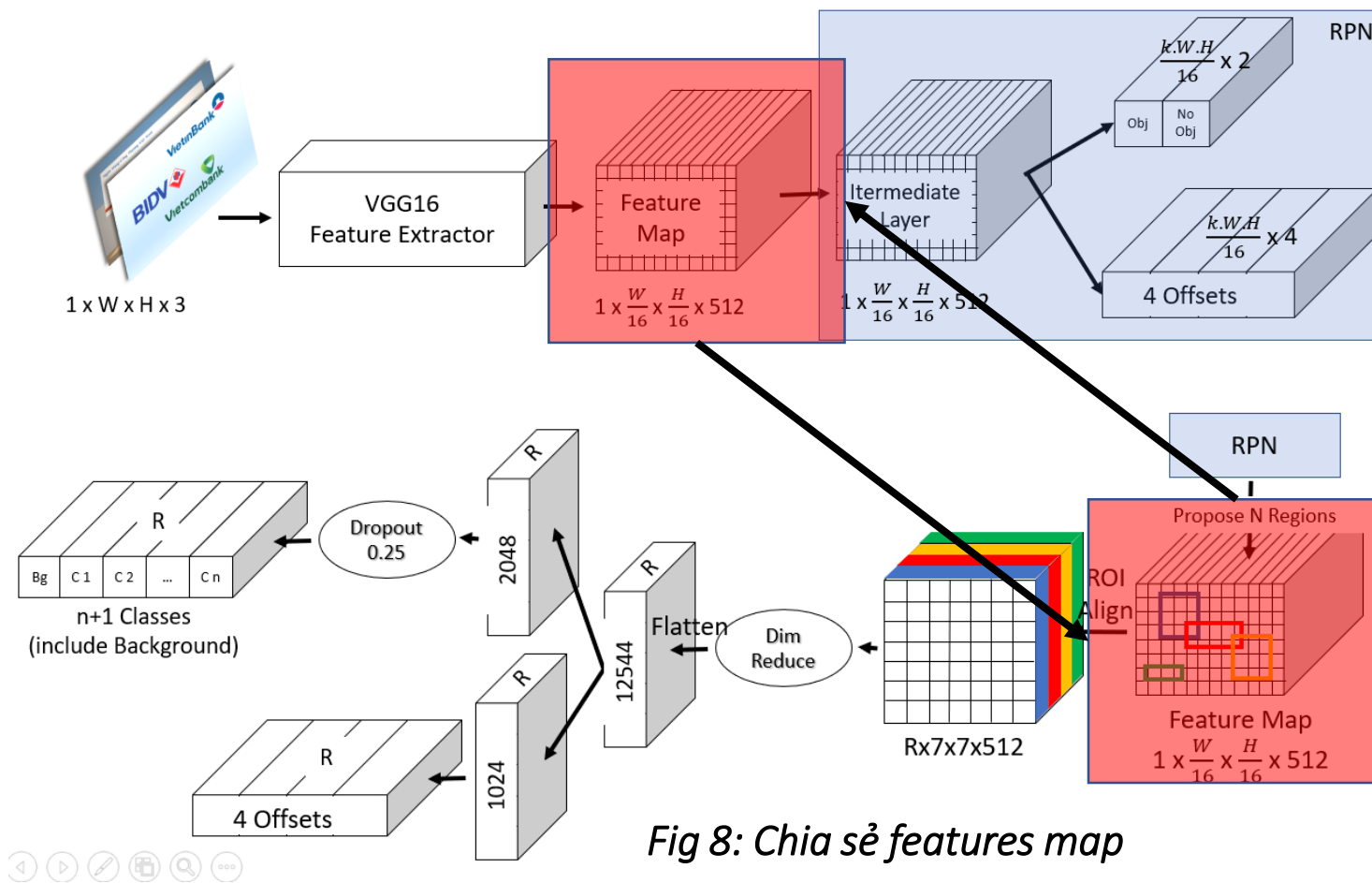


Fig 8: Chia sẻ features map

TỐI ƯU

- Sử dụng VGG16 để trích xuất đặc trưng & sử dụng pretrained VGG16 từ ImageNet

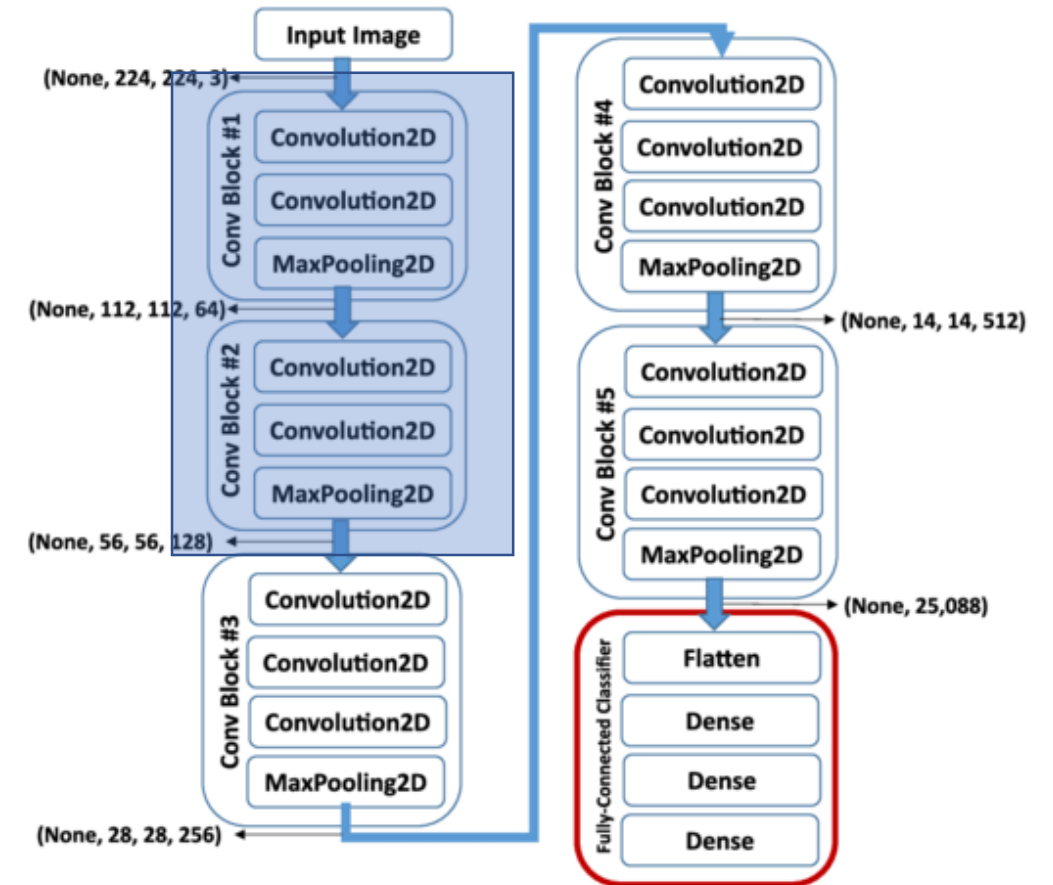


Fig 9: Đóng băng 2 khối đầu tiên trong mạng VGG16

TỐI ƯU

- Sử dụng VGG16 để trích xuất đặc trưng & sử dụng pretrained VGG16 từ ImageNet
- Tỷ lệ khung hình của ảnh được giữ nguyên có thể, kích thước tối thiểu 608 pixels, kích thước tối đa 1024 pixels



Fig 10: Thay đổi kích thước ảnh đầu vào

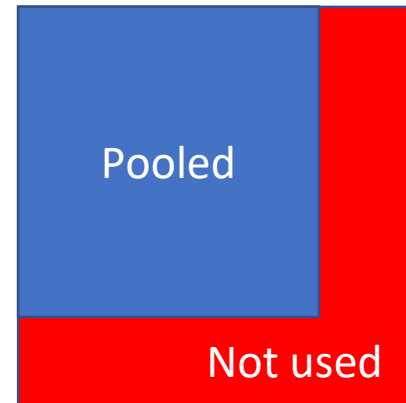
TỐI ƯU

- Sử dụng VGG16 để trích xuất đặc trưng & sử dụng pretrained VGG16 từ ImageNet
- Tỷ lệ khung hình của ảnh được giữ nguyên có thể, kích thước tối thiểu 608 pixels, kích thước tối đa 1024 pixels
- Sử dụng giải thuật NMS và thay thế ROI Pooling bằng ROI Align

- Đề xuất vùng kích thước 25 x 25
- Mục tiêu đưa về kích thước 7 x 7

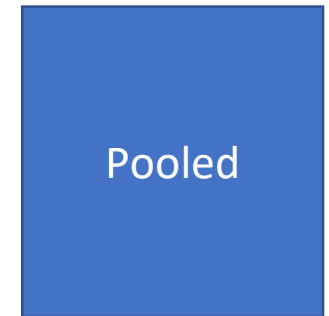
ROI Pooling

- Ta có $\lceil 25/7 \rceil = 4$
- Chỉ sử dụng 21x21



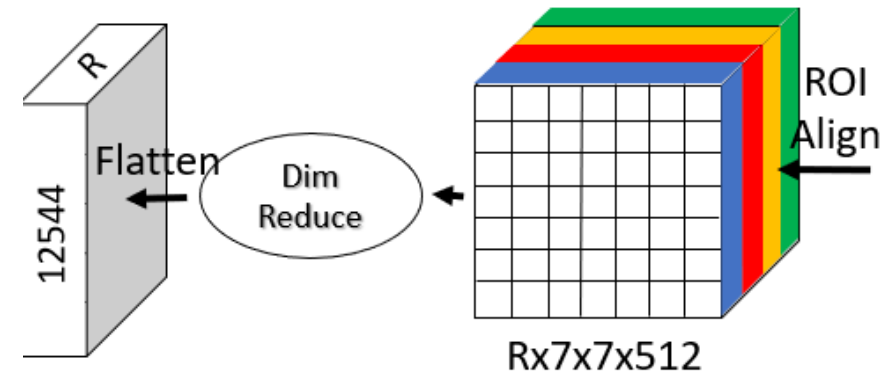
ROI Align

- Resize ảnh từ 25x25 về 14x14
- Thực hiện MaxPooling



TỐI ƯU

- Sử dụng VGG16 để trích xuất đặc trưng & sử dụng pretrained VGG16 từ ImageNet
- Tỷ lệ khung hình của ảnh được giữ nguyên có thể, kích thước tối thiểu 608 pixels, kích thước tối đa 1024 pixels
- Sử dụng giải thuật NMS và thay thế ROI Pooling bằng ROI Align
- Bổ xung lớp giảm số chiều dữ liệu sau lớp ROI Align



- Mặc định: $7 \times 7 \times 512 \times 2048 = 51M$
- Giảm số chiều: $7 \times 7 \times 256 \times 2048 + 1 \times 1 \times 256 \times 512 = 25.5M$

TỐI ƯU

- Sử dụng VGG16 để trích xuất đặc trưng & sử dụng pretrained VGG16 từ ImageNet
- Tỷ lệ khung hình của ảnh được giữ nguyên có thể, kích thước tối thiểu 608 pixels, kích thước tối đa 1024 pixels
- Sử dụng giải thuật NMS và thay thế ROI Pooling bằng ROI Align
- Bổ xung lớp giảm số chiều dữ liệu sau lớp ROI Align
- Giảm số chiều lớp hồi quy

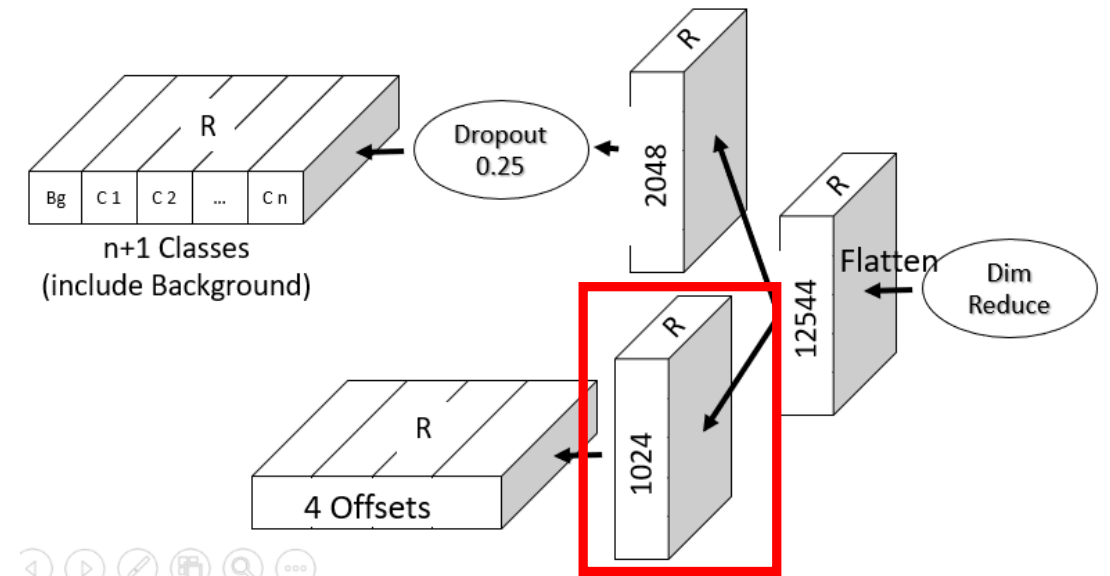


Fig 11: Giảm số chiều lớp hồi quy

TỐI ƯU

- Sử dụng VGG16 để trích xuất đặc trưng & sử dụng pretrained VGG16 từ ImageNet
- Tỷ lệ khung hình của ảnh được giữ nguyên có thể, kích thước tối thiểu 608 pixels, kích thước tối đa 1024 pixels
- Sử dụng giải thuật NMS và thay thế ROI Pooling bằng ROI Align
- Bổ xung lớp giảm số chiều dữ liệu sau lớp ROI Align
- Giảm số chiều lớp hồi quy
- Chỉ sử dụng dropout 25% cho lớp phân loại

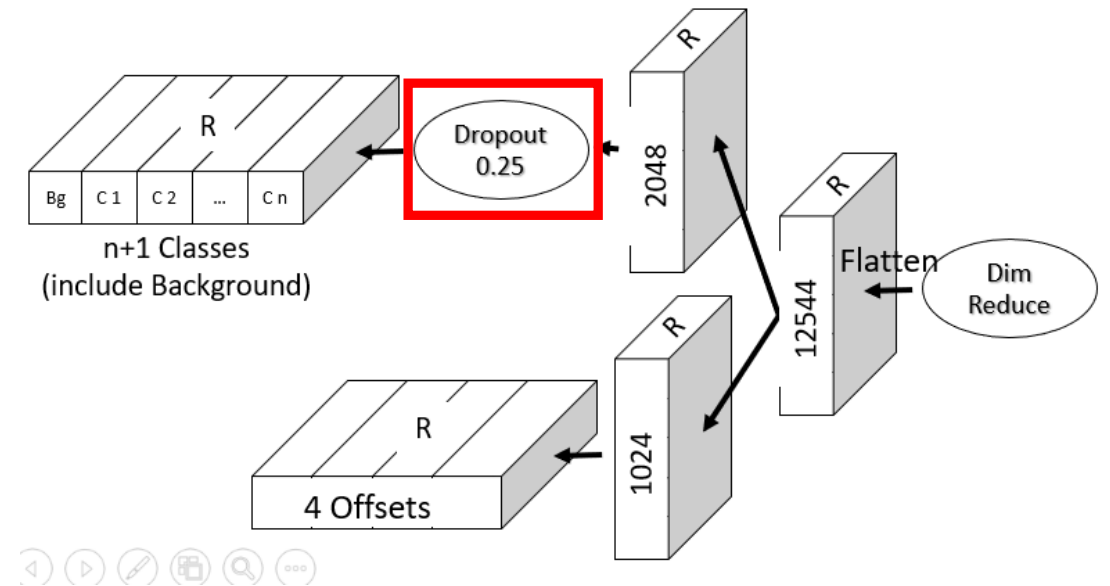


Fig 12: Dropout lớp phân loại, giảm Overfitting

TỐI ƯU

- Sử dụng VGG16 để trích xuất đặc trưng & sử dụng pretrained VGG16 từ ImageNet
- Tỷ lệ khung hình của ảnh được giữ nguyên có thể, kích thước tối thiểu 608 pixels, kích thước tối đa 1024 pixels
- Sử dụng giải thuật NMS và thay thế ROI Pooling bằng ROI Align
- Bổ xung lớp giảm số chiều dữ liệu sau lớp ROI Align
- Giảm số chiều lớp hồi quy
- Chỉ sử dụng dropout 25% cho lớp phân loại
- Loại bỏ anchors vượt quá kích thước ảnh

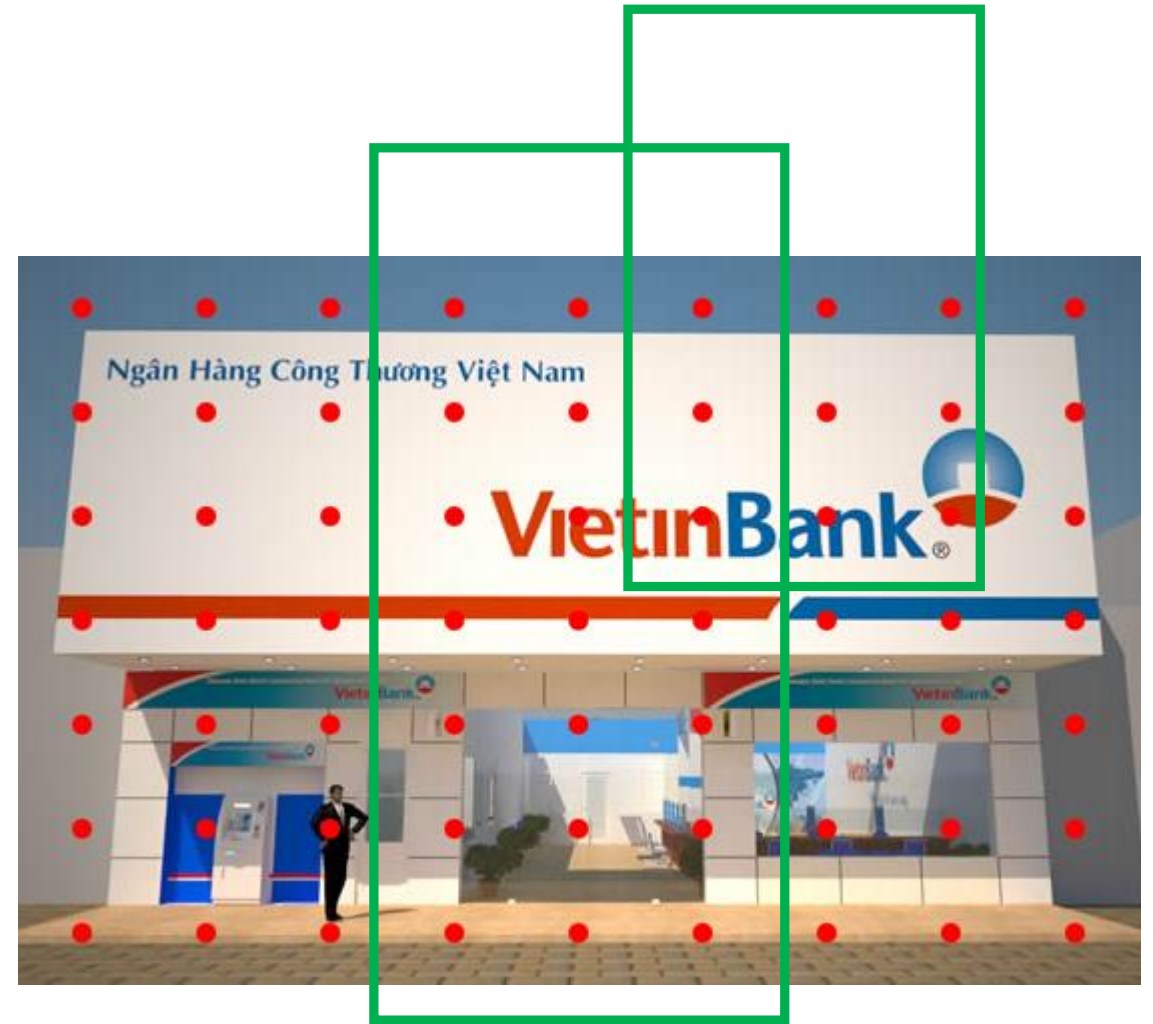


Fig 13: Loại bỏ anchors không cần thiết

KẾT QUẢ

- Độ chính xác tính trên tập 150 ảnh test
- Thời gian tính toán trung bình trên GTX 950M 4Gb của mô hình là **0.83s**, khi sử dụng mô hình gốc từ Tensorflow API là **1.67s**, nhanh hơn **2.01** lần

	VietinBank	VietcomBank	BIDV
False Positive	3	2	3
False Negative	5	3	4
True Positive	52	47	47
True Negative	95	101	100

	VietinBank	VietcomBank	BIDV
Precision	0.945	0.959	0.940
Recall	0.912	0.94	0.921

Fig 14: Độ chính xác mô hình trên tập test

KẾT QUẢ



Fig 15: Một số kết quả

KẾT QUẢ



Fig 15: Một số kết quả

XÂY DỰNG ỨNG DỤNG

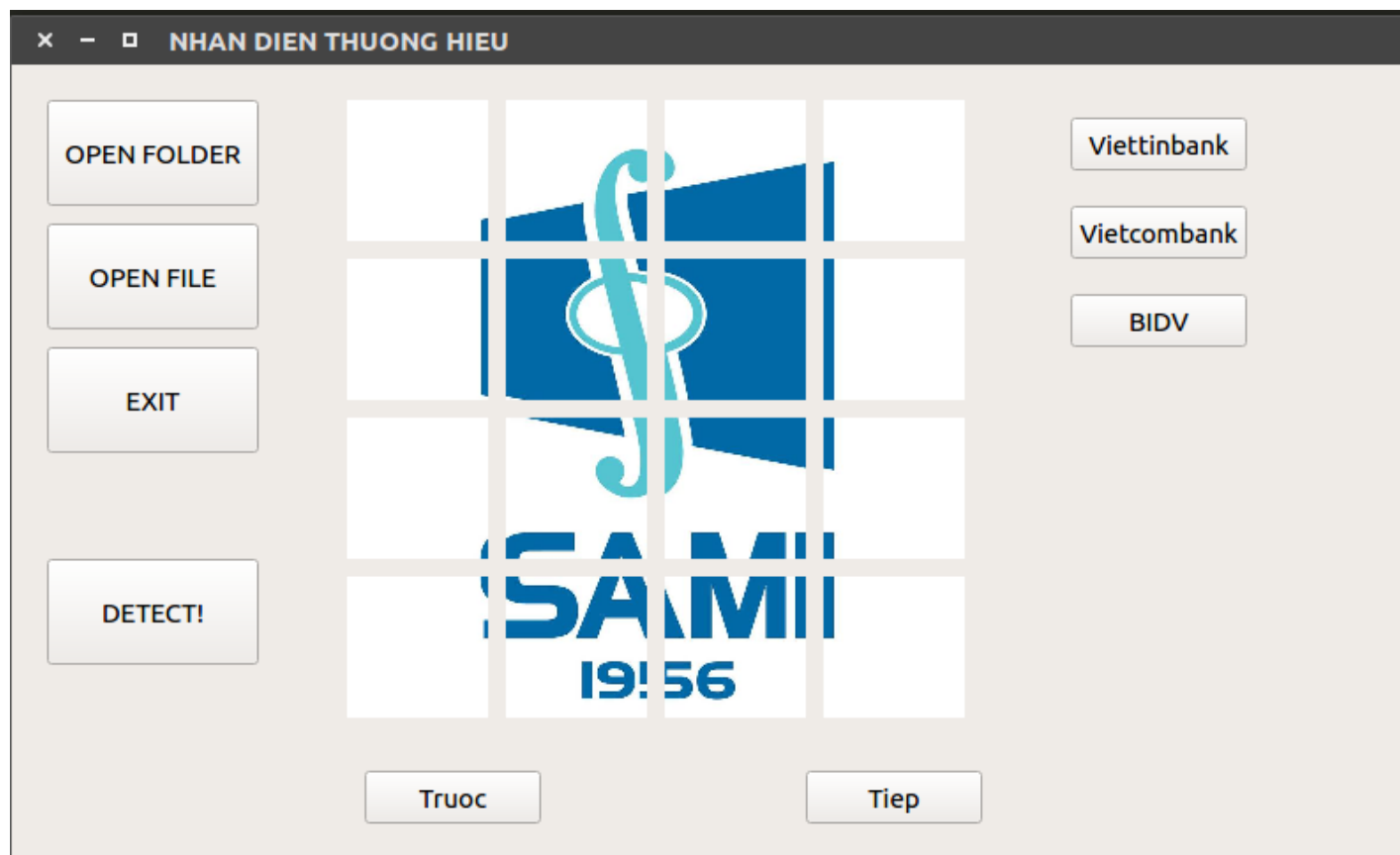


Fig 16: Xây dựng giao diện ứng dụng

XÂY DỰNG ỨNG DỤNG

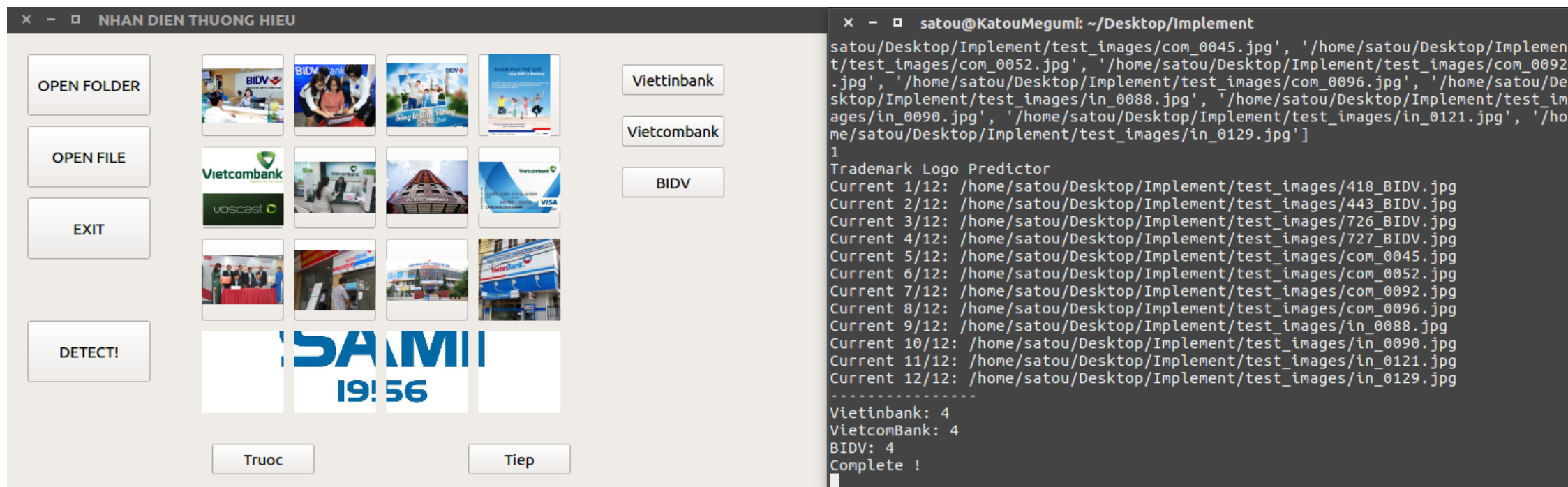


Fig 16: Xây dựng giao diện ứng dụng

TÀI LIỆU THAM KHẢO

- [1] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks - Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun.
- [2] SSD: Single Shot MultiBox Detector - Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg
- [3] You Only Look Once: Unified, Real-Time Object Detection- Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi.
- [4] VGG16, VGG19: Convolution Neural Network- Google Inc., 2015.
- [5] Rapid Object Detection using a Boosted Cascade of Simple Features - Paul Viola, Michael Jones, Computer Vision and Pattern Recognition 2001.
- [6] Scale-invariant feature transform (SIFT)- David Lowe, University of British Columbia, 1999
- [7] Lenet-5: Convolution neural network for handwritten digits recognition, Yann Lecun, 1998
- [8] Selective Search for Object Recognition - Jasper R. R. Uijlings, Koen E. A. van de Sande, Theo Gevers, Arnold W. M. Smeulders- International Journal of Computer Vision, Volume 104 (2), page 154-171, 2013
- [9] Efficient Graph-Based Image Segmentation - P. Felzenszwalb, D. Huttenlocher - International Journal of Computer Vision, Vol. 59, No. 2, September 2004
- [10] Rich feature hierarchies for accurate object detection and semantic segmentation - Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik- UC Berkeley- 22 Oct 2014
- [11] Deep Residual Learning for Image Recognition - Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Microsoft Research, 2015

TÀI LIỆU THAM KHẢO

- [12] Rethinking the Inception Architecture for Computer Vision - Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Google Inc, 2015
- [13] Xception: Deep Learning with Depthwise Separable Convolutions - Francois Chollet, Google Inc, 2017
- [14] Understanding the Effective Receptive Field in Deep Convolutional Neural Networks - Wenjie Luo*, Yujia Li*, Raquel Urtasun, Richard Zemel, Department of Computer Science, University of Toronto, 2017
- [15] Mask-RCNN - Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, Facebook AI Research, 24 2018
- [16] Panoptic-FPN - Alexander Kirillov, Ross Girshick, Kaiming He, Piotr Dollar, Facebook AI Research, 2019
- [17] SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5Mb model size - Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer
- [18] Non Maximal Suppression in Cascaded Ranking Models - Blaschko, M.B., Kannala, J., Rahtu, Scandinavian Conference on Image Analysis (SCIA). (2013)
- [19] Adam: A Method for Stochastic Optimization - Diederik P. Kingma*, Jimmy Lei Ba*, Jan 2017.
- [20] Making large-scale support vector machine learning practical -T. Joachims, . In: Schlkopf, B., Burges, C., Smola, A. (eds.) Advances in Kernel Methods – Support Vector Learning, pp. 169-184. MIT Press, Cambridge 1998.
- [21] A Survey on Deep Transfer Learning - Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu, Aug 2018