

**Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science**



Performance Evaluation of Pixel and Blocks-based Semantic Segmentation Method Using Artificial Neural Networks Models

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of the Requirements
for the Degree of Master of Science
in
Computer Science**

**Submitted by
Raya Akram Hamdy**

**Supervised by
Asst. Prof. Dr. Mohammed Chachan Younis**

2023 A.D.

1445 A.H.

ABSTRACT

Semantic segmentation is a crucial task in the machine learning domain that involves assigning a unique class label to each pixel in an image. While this pixel-by-pixel segmentation technique produces high-quality results, training and evaluating a semantic segmentation model can be a time-consuming process. To address this challenge, this study proposes a simplified version of semantic segmentation that involves block by block (5x5 Chunks) image fragmentation, feature extraction, and model training using specific color and textural features. To evaluate the proposed method, multiple images have been trained and tested using both pixel-by-pixel classification and the proposed block-based segmentation method. The results demonstrate that the proposed technique i.e., block-by-block segmentation is faster and more accurate than the traditional pixel-by-pixel classification method.

This thesis has utilized the Oxford-IIIT Pet dataset, from which thirty images of cats have been selected. Five different Artificial Neural Network (ANN) models were evaluated and trained, i.e., Levenberg-Marquardt Back Propagation (LM model), Broyden Fletcher Goldfarb Shanno (BFGS model), Resilient Back Propagation (RP model), Scaled Conjugate Gradient Back Propagation (SCG model), and Variable Learning Rate Back Propagation (GDx model) for both pixel-based and block-based methods. Pixel-based accuracy ranges from 70.82% to 76.47%, while block-based accuracy ranges from 82.94% to 85.83%. It is observed that the accuracy of all models improves in the block-based approach compared to the pixel-based approach. The processing time of each model for both pixel-based and block-based methods is also noted for both types. For the pixel-based approach, RP model takes the longest processing time

i.e., 242.39 seconds, while GDX model takes the shortest processing time i.e., 49.89 seconds. For the block-based approach, LM model takes the longest processing time i.e., 13.86 seconds, while GDX still has the shortest processing time i.e., 5.98 seconds. Therefore, block-based methods can be seen as more efficient and accurate for classification models.

An additional evaluation has been used to evaluate and compare the performance of the five models on three sample test images. These three images have been provided as an input to the trained models, which have segmented the cat object from the background. The accuracy and effectiveness of the segmentation method have been evaluated by comparing the predicted images generated by the trained models with the ground truth images. Results demonstrate that the accuracy of the LM model is the highest across all three test images, with 94.72%, 84.92%, and 89.81% for test images 1, 2, and 3, respectively. The GDX model has the lowest accuracy, with 92.96%, 66.9%, and 81.15% for test images 1, 2, and 3, respectively. The remaining models, RP, SCG, and BFG, have intermediate levels of accuracy.

A 64-bit Windows 10 OS, with an Intel(R) Core i7 processor (2.20GHz) with 16GB of RAM was used for the experimentation. The classification was achieved by using the tools available in the Matlab R2021b Machine Learning Toolbox.



وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

تقييم أداء طريقة التجزئة الدلالية القائمة على البكسل والكتل باستخدام نماذج مختلفة للشبكات العصبية الاصطناعية

رسالة مقدمة

الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
علوم الحاسوب

من قبل
ريا أكرم حمدي

بإشراف
أ.م.د. محمد جاجان يونس

الخلاصة

يُعد التقسيم الدلالي (التجزئة الدلالية) ذو مكانة بالغة الأهمية في مجال التعلم الآلي، إذ يتم في التقسيم الدلالي للبكسل الواحد تصنيف كل بكسل كوحدة واحدة (5*5) في الصورة لوصف محتوى ذلك البكسل. وعلى الرغم من أن تقنية تجزئة كل بكسل على حدا تنتج نتائج عالية الجودة، إلا أن عملية التدريب والتقييم في التجزئة الدلالية قد تستغرق وقتاً طويلاً. وللتغلب على هذه المشكلة، بسطت هذه الدراسة عملية التجزئة الدلالية من خلال تجزئة الصورة الى أجزاء أصغر تسمى الكتلة، ثم تم استخراج الميزات أو الخصائص الرئيسية وتدريب النموذج على تلك الميزات باستخدام خصائص الالوان والقوام لتلك الاجزاء. دُرِبَ العديد من الصور وأُختبرت باستخدام طريقة تصنيف كل بكسل على حدا وطريقة التجزئة المقترحة القائمة على الكتلة من خلال الخوارزمية المقترحة. وقد أوضحت النتائج أن تقنية التجزئة المقترحة، هي الأسرع والأكثر دقة من طريقة التصنيف التقليدية لكل بكسل على حدا.

استخدمت هذه الدراسة قاعدة بيانات Oxford-IIIT Pet، وأُختبرت ثلاثين صورة للقطط من مجموعة البيانات هذه. دُرِبَ وقُيِّمَت خمسة نماذج مختلفة للشبكات العصبية الاصطناعية لكل من الأسلوبين القائمين على البكسل وعلى الكتلة: نموذج الانتشار العكسي لفنبرك ماركورت، نموذج برويدن فليجر كولدفرب شانو، نموذج الانتشار الخلفي المرن، نموذج الانتشار الخلفي المتدرج المتقارب ونموذج الانتشار الرجعي لمعدل التعلم المتغير. أظهرت النتائج أن الدقة المستندة إلى البكسل تتراوح من 70.82% إلى 76.47%، بينما تتراوح الدقة القائمة على الكتلة من 82.94% إلى 85.83% وهذا يوضح أن دقة جميع النماذج تحسن في النهج القائم على الكتلة مقارنة بالنهج القائم على البكسل. ويُلاحظ أيضاً أن وقت المعالجة للنهج المستند إلى البكسل يستغرق وقت أطول من الأسلوب المستند إلى الكتلة، إذ يستغرق نموذج الانتشار الخلفي المرن للنهج القائم على البكسل 242.39 ثانية، بينما يستغرق نموذج الانتشار الرجعي لمعدل التعلم المتغير أقصر وقت معالجة أي 49.89 ثانية. أما بالنسبة للنهج القائم على الكتلة، فيستغرق

نموذج الانتشار العكسي لفنبرك ماركورت أطول وقت معالجة 13.86 ثانية، في حين لا يزال نموذج الانتشار الرجعي لمعدل التعلم المتغير أقصر وقت معالجة كما هو الحال في نهج البكسل 5.98 ثانية. لذلك، تُعد الطرائق القائمة على الكتلة أكثر كفاءة فيما يتعلق بالوقت والدقة لنماذج التصنيف.

وقد قُيِّم وقورن أداء نماذج الشبكة العصبية الخمسة المستخدمة في هذه الدراسة على عينة تتكون من ثلاث صور اختبار كمدخلات للنماذج المدربة مسبقاً والتي قَسِّمت كائن القط وخلفية الصورة. وقد قُيِّمت دقة وفعالية طريقة التجزئة الدلالية من خلال مقارنة الصور المتوقعة الناتجة عن النماذج المدربة مع صور الحقيقة الأرضية. وقد أوضحت النتائج أن دقة نموذج LM هي الأعلى لصور الاختبار الثلاثة 94.72% و 84.92% و 89.81% لصور الاختبار 1 و 2 و 3 على التوالي. أما نموذج الانتشار الرجعي لمعدل التعلم المتغير فقد أعطى أقل دقة حيث بلغت 92.96% و 66.9% و 81.15% لصور الاختبار 1 و 2 و 3 على التوالي. وبالنسبة للنماذج المتبقية فقد أعطت مستويات متوسطة من الدقة.

تم في هذه الرسالة استخدام نظام التشغيل Windows 10 64 بت، مع معالج Corei7 Intel(R) (2.20 جيجا هرتز) مع 16 جيجابايت من ذاكرة الوصول العشوائي. تم تحقيق التصنيف باستخدام الأدوات المتوفرة في صندوق تعلم الآلة الذي توفره لغة الماتلاب الإصدار 2021b المستخدمة في هذه الدراسة.