

Studying the Effect of Drop_Out Rate on the Performance of U_Net for Semantic Segmentation Image

Dr. Ghadi Mahmoudi*

Dr. Thanaa Jbeily**

Farah Haddad***

(Received 4 / 3 / 2022. Accepted 22 / 8 / 2022)

□ ABSTRACT □

Semantic segmentation in satellite images is a detailed technique to take advantage of space imaging, Deep learning networks are currently considered one of the basic techniques for solving many real-world problems, including processing satellite images, especially Encoder-Decoder Architecture. Designing the architecture of a deep learning network requires searching in a wide range of solutions, ranging from simple structures, which give modest results, to more complex structures, in the hope of reaching better results. These solutions come with different time and technology costs (the hardware on which the software runs).

This research attempts to investigate solutions corresponding to local conditions from the number of available satellite images and processing the raw images before entering them into the semantic segmentation stage, For implementing this stage, the research uses the modified U_Net network in terms of the size of the input layer, the number of convolutional layers, the type of activation function, the type of pixel Up_Sample process in the second path of the network, Moreover, research study's adding Drop_Out layers to the network structure (studying different percentages of Drop_Out and its effect on the model results).

Keywords: Deep Learning Networks, Semantic Segmentation, U_Net Network, Satellite Images, Drop-Out Rate.

*Assistant Professor, Department of communication and Electronics , Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria, g.mahmoudi@tishreen.edu.sy

**Assistant Professor, Department of communication and Electronics , Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria, thanaa.jbeily@tishreen.edu.sy

***Postgraduate Student (Master), Department of communication and Electronics, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria, farah.haddad@tishreen.edu.sy

دراسة أثر نسبة التسرب على أداء شبكة U_Net للتقسيم الدلالي للصور

د. غادي محمودي*

د. ثناء جبيلي**

فرح حداد***

(تاريخ الإيداع 4 / 3 / 2022. قُبِلَ للنشر في 22 / 8 / 2022)

□ ملخص □

يعتبر الفصل الدلالي في الصور الفضائية تقنية مفصلية للاستفادة من التصوير الفضائي وتعد شبكات التعلم العميق حالياً من التقنيات الأساسية لحل الكثير من المسائل الواقعية، ومنها معالجة الصور الفضائية، وخصوصاً بنية مرمر_فاك ترميز (Encoder-Decoder Architecture). يتطلب تصميم بنية شبكة التعلم العميق البحث في فضاء واسع من الحلول، تتراوح بين البنى البسيطة، التي تعطي نتائج متواضعة، والبنى الأكثر تعقيداً، طمعا بالوصول إلى نتائج أفضل. ترافق هذه الحلول كلف زمنية وتقنية (الأجهزة التي تعمل عليها البرمجيات) مختلفة. يحاول هذا البحث استقصاء الحلول الموافقة للظروف المحلية من عدد الصور الفضائية المتاحة ومعالجة الصور الأولية قبل إدخالها إلى مرحلة التقسيم الدلالي والتي سيتم تنفيذها باستخدام شبكة U_Net المعدلة من حيث حجم طبقة الإدخال وعدد الطبقات الالتفافية ونوع دالة التنشيط ونوع عملية استعادة البكسلات في المسار الثاني من الشبكة بالإضافة إلى دراسة تأثير إضافة طبقات تسرب إلى بنية الشبكة (دراسة نسب مختلفة للتسرب وتأثيره على نتائج النموذج).

الكلمات المفتاحية: شبكات التعلم العميق، الفصل الدلالي، شبكة U_Net، الصور الفضائية، نسبة التسرب.

* مدرس-قسم هندسة الاتصالات والالكترونيات - كلية الهندسة الميكانيكية والكهربائية-جامعة تشرين - اللاذقية- سورية. بريد الكتروني: g.mahmoudi@tishreen.edu.sy
**مدرس-قسم هندسة الاتصالات والالكترونيات - كلية الهندسة الميكانيكية والكهربائية-جامعة تشرين - اللاذقية- سورية. بريد الكتروني: thanaa.jbeily@tishreen.edu.sy
***طالبة دراسات عليا(ماجستير)-قسم هندسة الاتصالات والالكترونيات-كلية الهندسة الميكانيكية والكهربائية-جامعة تشرين-اللاذقية- سورية. بريد الكتروني: farah.haddad@tishreen.edu.sy

مقدمة:

تتميز الصور الجوية (الفضائية) بغزارة محتوى المعلومات وشموليتها، مما ينعكس على الحجم الكبير للصورة وما يترتب عليه ذلك من مراعاة لخصائصها في تصميم النموذج. اقترحت العديد من الأبحاث نماذج شبكات تعلم عميق لمعالجة الصور الجوية (الفضائية) سواء على مستوى التصنيف Classification، التقسيم الدلالي Semantic Segmentation، والتقسيم اللحظي Instance Segmentation، ولكن نتيجة اختزال حجم الصورة أثناء عملية المعالجة وبالتالي فقد كمية كبيرة من المعلومات المكانية ظهرت مشاكل متعددة في الصور المعالجة النهائية [1,2,3]. بهدف حل هذه المشاكل تم اقتراح بنية المرمز_فاك الترميز (Encoder_Decoder Architecture) لتطبيق التقسيم الدلالي على الصور الفضائية باستخدام قاعدة بيانات كبيرة، أدى هذا الاقتراح إلى تخفيض فقدان المعلومات المكانية، وتحسين كفاءة الشبكة في التصنيف على مستوى البكسل، إلا أنه زاد التعقيد الحسابي والزمن اللازم لإنهاء مرحلتي التدريب والاختبار [4,5].

أهمية البحث وأهدافه:

تعد شبكات التعلم العميق من التقنيات الأساسية لحل الكثير من المسائل الواقعية، ومنها معالجة الصور الفضائية، مع الإشارة إلى أن جامعة تشرين لديها أبحاث في مجال إطلاق قمر صناعي من المتوقع أن يكون قادراً على إرسال صور تتطلب المعالجة للاستفادة منها في العديد من التطبيقات. يهدف البحث إلى تطبيق التقسيم الدلالي على قاعدة بيانات صغيرة متاحة للأغراض البحثية بغرض إيجاد بنية متوسطة الكلف الزمنية والتقنية، وبالتالي القدرة على اختبار فعالية هذه البنية على الصور التي قد تصبح متاحة محلياً.

طرائق البحث ومواده:

يدرس هذا البحث بنية شبكة تعلم عميق من أجل التقسيم الدلالي للصور الفضائية، وأجريت المحاكاة على برنامج Spyder Version 5 برنامج محاكاة مفتوح المصدر يعمل على نظام Windows وقد تم استخدامه في أبحاث سابقة.

1. الدراسات المرجعية:

اهتمت العديد من الدراسات بتطوير شبكات التعلم العميق الالتفافية بغرض تصنيف أكثر دقة على مستوى البكسل في الصور العادية، الطبية، الفضائية.

اقترحت بعض الدراسات تدريب شبكات على صور عادية أو طبية رغم قلة البيانات وتحديد موطن البكسل بدقة في الصور الطبية [6]، بالإضافة إلى تقليل حجم الذاكرة المستخدم والتعقيد [9]، استخدام أكثر من بنية مرمر AlexNet، VGG net & GoogLeNet في (Fully Connected Network)FCN، بالإضافة إلى استخدام وصلات اتصال التخطي (Skip Connection) لنقل المعلومات المكانية من أجل التجزئة الدلالية للصور العادية [10]، المقارنة بين قدرة الشبكات العصبية العميقة وقدرة العقل البشري على التجزئة الدلالية للبكسلات في الصورة العادية عند تدهور الصورة، وبالتالي أظهرت أهمية التطوير والتعديل المستمر على الشبكات العصبية العميقة المستخدمة للتقسيم الدلالي [9].

تحليل صور الأقمار الصناعية في العديد من الدراسات: من أجل الكشف المبكر عن الكوارث ومراقبة البيئة [10]، تصميم إطار آلي لاستخراج الخرائط الدلالية للطرق والطرق السريعة لتتبع المناطق الحضرية ونمو المدن باستخدام

Modified FCN Architecture (بنية الشبكات العصبونية كاملة الاتصال المعدلة) [11]، استخراج الميزات من الصور الفضائية وتصنيفها باستخدام شبكات CNN (الشبكات العصبونية الالتفافية Convolutional Neural Network) لتحسين طرق الكشف عن المركبات [12]، استخدام شبكات U-Net وتابع Lovasz- Softmax loss لتصنيف الغطاء الأرضي [13]، تصنيف أراضي المراعي في منطقة وسط البرازيل باستخدام U-NET & LSTM (Long Short Term Memory) & random forest من صور أخذت من القمر الصناعي PlanetScope على مدى 12 شهر [14]، بعض الأبحاث أخذت منحى لتطبيقات أكثر حساسية مثل تطوير أنظمة اكتشاف الحرائق باستخدام الصور الفضائية وبيانات الطقس بحيث تكون المعلومات متزامنة، باستخدام CNN & LSTM وتزويدها بما يلي: قاعدة بيانات، توقعات الطقس، وأي تغير يحدث في الجو من ارتفاع درجة الحرارة انخفاض الرطوبة تغير ضغط الهواء وذلك عن طريق المعلومات الواردة من محطات التنبؤ الجوي القريبة [15]، وفي أبحاث أخرى تم التركيز على اكتشاف الدخان باستخدام صور الأقمار الصناعية من أجل اكتشاف حرائق الغابات وذلك باستخدام شبكات CNN & Channel-Wise Attention [16].

ومع التقدم في مجال تحليل الصور الفضائية وأهميته تم تعديل بنية U-Net لتصنيف الغطاء الأرضي وهي شبكات U-Net معدلة من حيث عدد الطبقات الالتفافية وحجم المرشحات وعدد المعاملات المدخلة، حققت هذه الدراسة سرعة في الكشف وتقليل في التعقيد مقارنة بباقي الشبكات FCN & U-Net [18].

درس البحث البنية الأكثر كفاءة لتطبيق التقسيم الدلالي على قاعدة بيانات لصور فضائية قليلة العدد نظرا لصعوبة الحصول على قاعدة بيانات صور فضائية كبيرة في سوريا، وبالتالي أهمية تصميم نموذج قادر على التعامل مع قاعدة بيانات صغيرة، وذلك لأنه عند استخدام قاعدة بيانات صغيرة لتدريب النموذج يحدث ما يسمى فرط التخصص (Overfitting) عند اختبار النموذج. وذلك ما سيتم دراسة تقليله وفق عدة سيناريوهات بقييم متغيرة لنسبة التسرب ضمن بنية الشبكة لتحقيق أفضل النتائج.

2. بنية الشبكة للتقسيم الدلالي

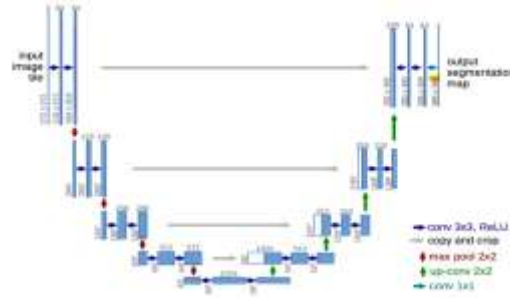
1.2. التقسيم الدلالي: يوجد العديد من المستويات لفهم الصورة حاسوبيا (الشكل (1))، بداية بتصنيف الكائن (Classification) الموجود بالصورة، ثم بتحديد موقعه (Localization)، واكتشاف الأشياء بالصورة (Object Detection)، ونتيجة التداخل الذي من الممكن أن يحدث بين البكسلات التابعة لأصناف مختلفة تم الانتقال إلى تصنيف على مستوى البكسل (Semantic Segmentation)، وفي بعض الأحيان يحدث تداخل بين البكسلات التابعة لنفس الصنف فاقترحت أيضا طريقة تصنيف للبكسلات التابعة لنفس الصنف (Instance Segmentation).



الشكل (1): مستويات فهم الصورة حاسوبيا

في التقسيم الدلالي المهمة الرئيسية هي إمكانية تحديد كل بكسل في الصورة إلى أي صنف تابع وتحليل بيانات التدريب من خلال خوارزميات التعلم العميق من أجل التحليل الصحيح والدقيق للمشهد، يوجد نوعين للتقسيم الدلالي: التقسيم الدلالي الثنائي (Binary Semantic Segmentation) المقصود به تصنيف البكسلات إلى صنفين فقط (الصنف المطلوب والخلفية)، التقسيم الدلالي المتعدد الأصناف (Multiclass Semantic Segmentation) المقصود به تصنيف البكسلات إلى أكثر من صنف.

2.2. شبكة U_Net:



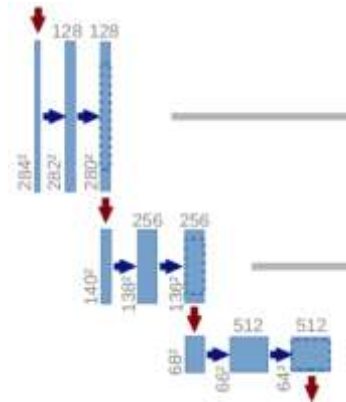
الشكل (2): U-Net Architecture

تم استخدام شبكات U-Net (الشكل (2)) التي اقترحت للمرة الأولى من أجل تحسين التقسيم الدلالي للصور الطبية الحيوية في عام 2015، من أجل التمكن من فصل الأشياء عن بعضها وعن الخلفية الخاصة بالصورة، وتبرز أهمية ذلك في صور الاستشعار عن بعد وفي صور الطب الحيوي حيث يكون من الضروري تقسيم الصور على مستوى البكسل بدقة، سميت شبكات U-Net لأنها تأخذ شكل حرف U تتكون من مسارين (الشكل (3)): المسار اليساري هو Contracting Path المكون من عدة عمليات التفاضلية، والمسار الثاني وهو المسار اليميني Expansive Path الذي سيقوم Up_Sampling Technic [6].



الشكل (3): U-Net Architecture

في المسار الأول Contracting Path: كما هو موضح في الشكل (4) كل عملية مكونة من طبقتين التفاضيتين، وعدد القنوات يتغير من 1 إلى 64، والسهم الأحمر يشير إلى عملية max pooling حيث تقلل حجم الصورة إلى النصف، أما اختلاف القياسات للصورة $568 \times 568 \rightarrow 572 \times 572$ نتيجة عملية padding تم استخدام padding = same، ثم تكرر العملية نفسها ثلاث مرات.



الشكل (4): Contracting Path U-Net Architecture

ولكن خلال تطبيق الطريقة الالتفافية للمرشحات على الصورة المدخلة ومن ثم تطبيق Pooling (عملية التجميع) حيث يتم تجميع العديد من الميزات بغاية تخفيف العمليات الحسابية فتبقى الميزات الأهم ضمن الخرج ولكن حجم الصورة سينخفض، كلما تم التعمق بالطبقات أكثر كلما استخرج ميزات وتصنيف أفضل ولكن فقدت القدرة على تحديد مكان البكسلات بدقة [6].

By down sampling, the model better understands “WHAT” is present in the image. But it loses the information of “WHERE” it is present.

إن ناتج تطبيق عملية Semantic Segmentation يجب أن يكون تصنيف على مستوى البكسل، ويجب أن تكون صورة الخرج عالية الدقة وبنفس حجم صورة الدخل وبالتالي الحاجة إلى Up- Sampling وهو المسار الثاني في شبكة U-Net ويمكن اعتبار العمليات التي تتم في هذا المسار هي عمليات متكاملة للمعلومات من الطبقات الحالية والطبقات السابقة.

2.3. التعديلات على شبكة U-Net:



الشكل (5): التعديلات على شبكة U-Net

اقترحت العديد من الدراسات تعديلات على بنية شبكة U-Net كما هو موضح في الشكل (5) بعضها يقترح التعديل على مستوى بنية الشبكة لتخفيف التعقيد وزيادة دقة الصور المقسمة دلاليًا بزيادة عدد القنوات التي تتم معالجتها أو استخدام وصلات التخطيطي بعدة سيناريوهات بغاية نقل المعلومات التي من الممكن أن تفقد أثناء عملية التقسيم الدلالي وتركيز

الانتباه على أوزان أكثر من أوزان لكي نحصل على صورة متوازنة ونتجنب ظهور رقع الشطرنج في الصورة المعالجة النهائية [9]. اقترحت أيضا تعديلات في مجال البنية الوظيفية لهذه الشبكة من خلال استخدام دوال وتوابع رياضية مختلفة، مثلا لاكتشاف الحواف يستخدم تابع التجميع الأعظمي وهذا ما سيختلف من أجل حالات تقسيم دلالي أخرى، واختلاف طريقة الالتفاف وحجم المرشحات وعددها بالإضافة إلى اختلاف الخطوة وبالتالي اختلاف عدد مرات الالتفاف على البكسل نفسه [10].

النتائج والمناقشة:

1. بنية الشبكة المستخدمة في البحث:

استخدم البحث Transpose Up_Sampling وهي عبارة عن عملية التناظرية (أي تحتوي على بارامترات محاكاة تعلم وعملية Up_Sampling)، وغيرنا حجم الإدخال عن بنية U_Net الأساسية من $572*572$ إلى $256*256$. واستخدم قاعدة بيانات لصور فضائية (جوية) لمدينة دبي [18] نظرا لعدم توفر قاعدة بيانات كافية في الهيئة العامة للاستشعار عن بعد. ونظرا لأهمية مشكلة قلة عدد الصور الفضائية المتوفرة محليا اتخذ البحث مسار تدريب النموذج على قاعدة بيانات صغيرة لكي يكون قادرا على إعطاء نتائج عندما تصبح الصور الفضائية متاحة محليا من المشاريع التي يتم دراستها وتطبيقها في جامعة تشرين (وهذه النقطة الأساسية المشتركة بين قاعدة البيانات المستخدمة في البحث) الخاصة بدبي) وقاعدة البيانات التي من الممكن أن تتوفر في سورية، بالإضافة إلى التأكيد على وجود صنف للبكسلات غير المصنفة وبالتالي التتويه إلى أنه من الممكن أن يكون هناك اختلاف بين قاعدة بيانات التدريب وقاعدة بيانات الاختبار). تتألف قاعدة البيانات من 72 صورة وقناع مقسمة إلى ملفات. تم التقسيم الدلالي للأقنعة المرافقة للصور من قبل منظمة رؤيا السورية إلى 6 أصناف مختلفة (ليتم إعطاء طابع باللون الكحلي للأبنية، أزرق مائل للبنفسجي للأراضي، أزرق فاتح للطرق، أصفر للغطاء النباتي، كموني للمياه، ورمادي للبكسلات غير القادر على تحديد الصنف التابعة له) مرمزة وفق النظام الست عشري:

Building: #3C1098

Land (unpaved area): #8429F6

Road: #6EC1E4

Vegetation: #FEDD3A

Water: #E2A929

Unlabeled: #9B9B9B

الصور ليست بنفس الأبعاد وإنما بأبعاد مختلفة مما يشكل تحدي إضافي في معالجة هذه الصور حيث سيتم اقتطاع (Patchify) هذه الصور لتصبح بحجم إدخال واحد وهو $256*256$ كما هو موضح في الجد

الجدول (1): جدول يظهر الأبعاد المختلفة للصور المستخدمة وكيفية معالجتها وتحويلها إلى حجم واحد 256*256

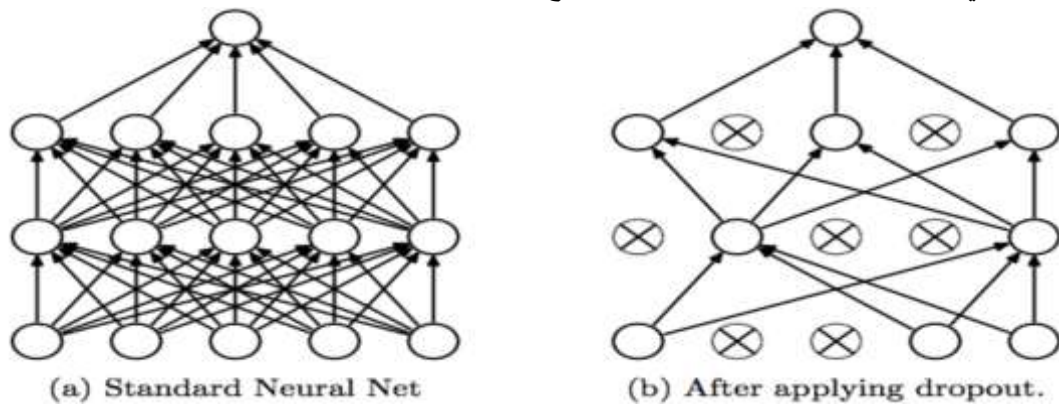
اسم الملف	أبعاد الصور في الملف	أبعاد الصور بعد التقريب إلى أقرب رقم صحيح للتقسيم إلى 256*256	عدد الصور المقطعة من الصورة الواحدة في كل ملف
Tile1	797 x 644	768 x 512	6
Tile2	509 x 544	512 x 256	2
Tile3	682 x 658	512 x 512	4
Tile4	1099 x 846	1024 x 768	12
Tile5	1126 x 1058	1024 x 1024	16
Tile6	859 x 838	768 x 768	9
Tile7	1817 x 2061	1792 x 2048	56
Tile8	2149 x 1479	1280 x 2048	40
Total 9 images in each folder * (145 patches) = 1305 Total 1305 patches of size 256x256			

ثم تحويل القيم الستة عشرية إلى قيم RGB ليتم تحويلها لاحقا إلى قيم صحيحة وتمثيلها ك أصناف Label، وأخيرا تقسيمها إلى بيانات تدريب وبيانات اختبار لنبدأ بعدها بتحديد بارامترات المحاكاة للنموذج. ضبطنا الأوزان الابتدائية للنموذج إلى 0.1666 كونها أوزان ابتدائية وسيتم التصنيف إلى 6 أصناف مختلفة، تم استخدام 9 طبقات التفاضلية تختلف بحجم وعدد المرشحات وباقي بارامترات المحاكاة الخاصة بكل طبقة تم ضبطها وفق التالي: (activation='relu', kernel_initializer='he_normal', padding='same')

أما (Activation Function دالة التنشيط) في طبقة الخرج واستخدام Softmax لكي يتم التصنيف إلى صنف محدد كنتيجة نهائية.

3. نسبة التسرب:

الشبكات العصبونية المدربة على قاعدة بيانات صغيرة نسبيا، عندما تدخل مرحلة الاختبار الحقيقية على صور جديدة تعاني من مشكلة التعميم نتيجة فرط التخصص الذي حدث في مرحلة التدريب على صور قليلة. التسرب هو طريقة تنظيم لتدريب عدد كبير من الشبكات العصبية ذات البنى المختلفة على التوازي، يتم افتراض قيم التسرب وليس حسابها، بحيث يمكن تعديلها وفقاً للنتائج. أثناء التدريب، يتم تجاهل عدد من مخرجات الطبقة بشكل عشوائي أو "يتم استبعادها" كما هو موضح في الشكل (6). هذا ما يجعل الطبقة تبدو وكأنها طبقة مع عدد مختلف من العقد والاتصالات بالطبقة السابقة. من خلال إسقاط وحدة ما، فهذا نعني إزالتها مؤقتاً من الشبكة، إلى جانب جميع اتصالاتها الواردة والصادرة. [9]



الشكل (6): Applying Drop-Out

يمكن أن يتم استخدام التسرب في كل أنواع الطبقات المخفية وينسب مختلفة وذلك في مرحلة التدريب، ولا يمكن استخدام التسرب في طبقة الخرج ولا في مرحلة الاختبار.

3. معاملات التقييم المدروسة:

3.1. تابع الخسارة (Function Loss):

يستخدم في شبكات التعلم العميق من أجل تقييم عمل خوارزمية الانحدار الاشتقاقي (Gradient Descent): المستخدمة من أجل تعديل الأوزان، فتبقى الأوزان تعدل عن طريق مسار الانتشار الخلفي حتى نحصل على أقل قيمة لتابع الخسارة [1].

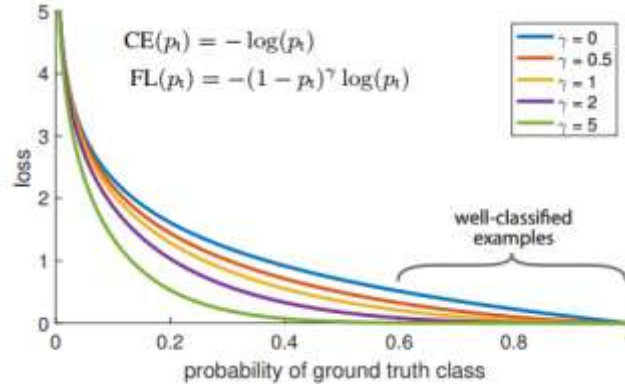
في التقسيم الدلالي الثنائي حقق تابع Dice_Loss أداء جيد نسبياً مع نسبة تداخل منخفضة بين البكسلات المطلوب تصنيفها وفصلها عن الخلفية وفق المعادلات الرياضية التالية:

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p} + 1}{y + \hat{p} + 1}$$

أما في حالة التقسيم الدلالي متعدد الأصناف فظهرت مشكلة الكلفة الزمنية العالية لاستخدام تابع الخسارة السابق نتيجة عدم القدرة على حسم قرار انتماء بكسل إلى صنف محدد بسبب التداخل الكبير بين الأصناف وكثرة عددها، فتم اقتراح استخدام تابع Focal_Loss وهو تابع يعمل بنفس طريقة تابع Cross_Entropy مع إضافة بارامتر جديد γ كما هو موضح في الشكل (7) وفق المعادلات الرياضية التالية:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

عندما $\gamma=1$ يعود عمل التابع كأنه Cross Entropy.



الشكل (7): Focal_Loss & Cross_Entropy

3.2. تابع نسبة التقاطع (Intersection Over Union):

وهي قيمة بين (0,1) تدل على التقاطع بين الحقيقة والتوقع. إذا كانت $IOU > 0.5$ ، فإنها تعتبر اكتشافاً صحيحاً، وإلا فإنها تعتبر اكتشافاً كاذباً [2].

الجدول (2): جدول يظهر رموز البكسلات الحقيقية ورموز البكسلات المتوقعة

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

3.3. الدقة (Accuracy):

اعتماداً على المعاملات في الجدول (2) يمكن تعريف الدقة على أنها نسبة البكسلات المصنفة بشكل صحيح إلى العدد الكلي للبكسلات (بغض النظر عن الصنف) [2].

$$\text{Overall Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

تم تدريب واختبار النموذج على 100 epochs وفق ثلاث سيناريوهات: (عند استخدام قاعدة بيانات صغيرة الحجم ظهور مشكلة فرط التخصص **Overfitting** في مرحلة الاختبار على صور جديدة، وعلى الرغم أنه يبدو أن عدد بارامترات المحاكاة في الشبكة الالتفافية قليل، إلا أنه يوجد **1943190 Parameter** في هذا النموذج مما يؤكد الحاجة إلى استخدام نسب تسرب واختبار تأثيرها على الشبكة).

في السيناريو الأول بدون استخدام التسرب، في السيناريو الثاني تم استخدام نسبة تسرب في كل طبقة 0.2 إلا في طبقة العنق بنسبة 0.3، أما السيناريو الثالث كان باستخدام نسب تسرب تصاعدياً (أي في الطبقة الالتفافية الأولى نسبة تسرب 0.1، في الطبقة التي تليها 0.2 وهكذا حتى نهاية مسار التشفير، ليتم عكس القيم في المسار الثاني) وملاحظة فرق النتائج والكلف الزمنية والتقنية.

أجريت المحاكاة على برنامج Spyder Versio 5 المنصب عن طريق Anaconda في بيئة مختلفة عن البيئة الافتراضية ب استخدام Tensorflow 2.7.0 & Keras 2.7.0 على نظام تشغيل Windows 10، استخدمت مكتبة Tensorflow & Keras.

بارامترات التقييم هي (تابع الخسارة، الدقة، نسبة التقاطع، نسبة التسرب، وقت التدريب، عدد عصور التدريب) وهي بارامترات التقييم الأكثر استخداماً في أبحاث التقسيم الدلالي على الصور الفضائية والصور الطبية أيضاً.

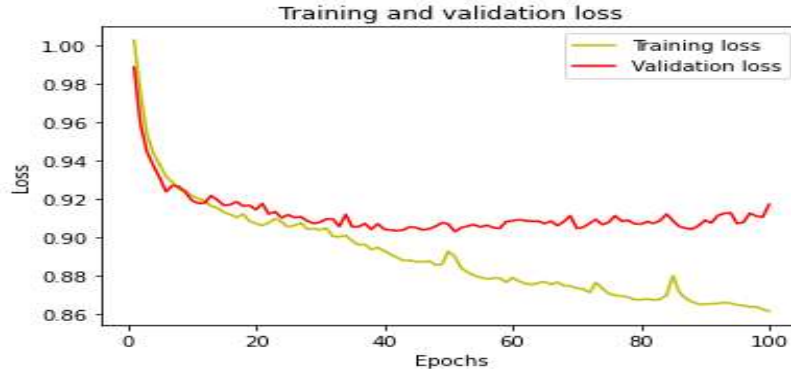
النتائج والمناقشة:

في الجدول (3) التالي قيم لتطبيق ثلاث سيناريوهات مختلفة للتدريب والاختبار، موضحاً اختلاف وقت التدريب وعدد عصور التدريب وقيم توابع الخسارة والدقة والتقاطع ومتوسط نسبة التقاطع وذلك في مرحلتي التدريب والاختبار.

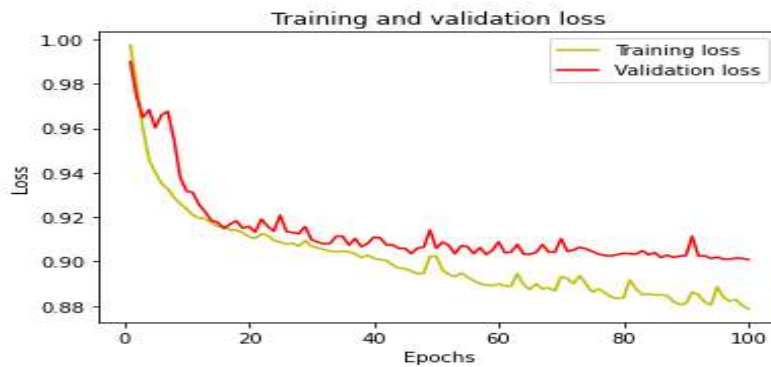
الجدول (3): جدول يظهر قيم معاملات الشبكة وفق ثلاث مراحل تدريب واختبار

Drop_out rate	Drop_out=0.0	Drop_out=0.2 at all steps Except at bottleneck =0.3	Drop_out Ascending order start with 0.1
Training Time	20 hours	32 hours	35 hours
Loss	0.86	0.87	0.88
Accuracy	0.95	0.90	0.89
Jacard-Coeff	0.88	0.79	0.76
Val-Loss	0.92	0.90	0.90
Val-Accuracy	0.84	0.87	0.87
Val-Jacard-Coeff	0.70	0.74	0.73
Mean IOU	0.58	0.62	0.61

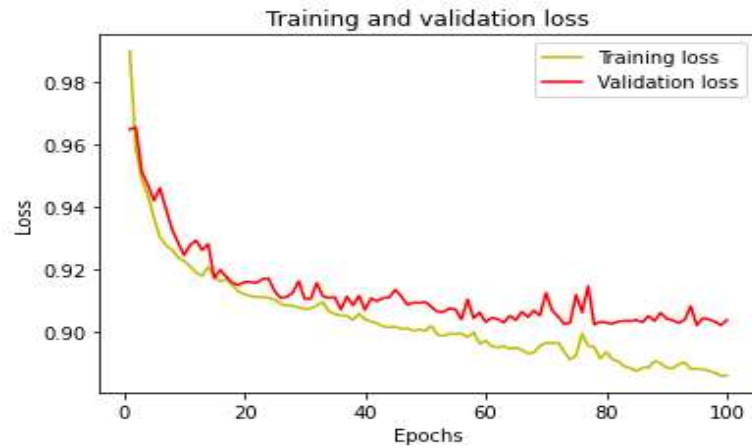
1. تابع الخسارة:



الشكل (8): تابع الخسارة للنموذج على 100 epochs بدون نسبة تسرب



الشكل (9): تابع الخسارة للنموذج على 100 epochs مع نسبة تسرب

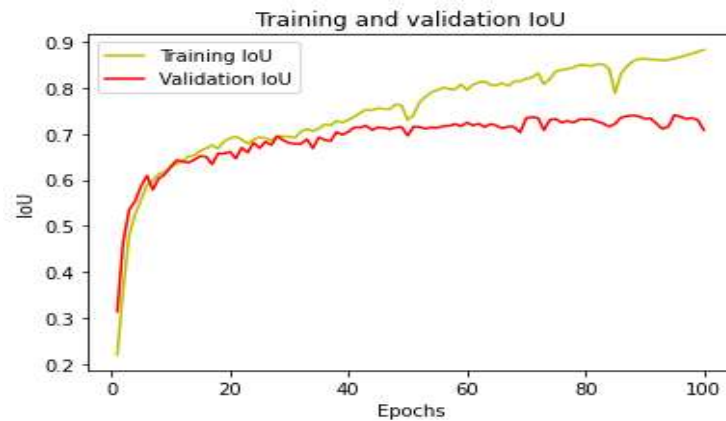


الشكل (10): تابع الخسارة للنموذج على 100 epochs مع تطبيق نسبة تسرب تصاعديا

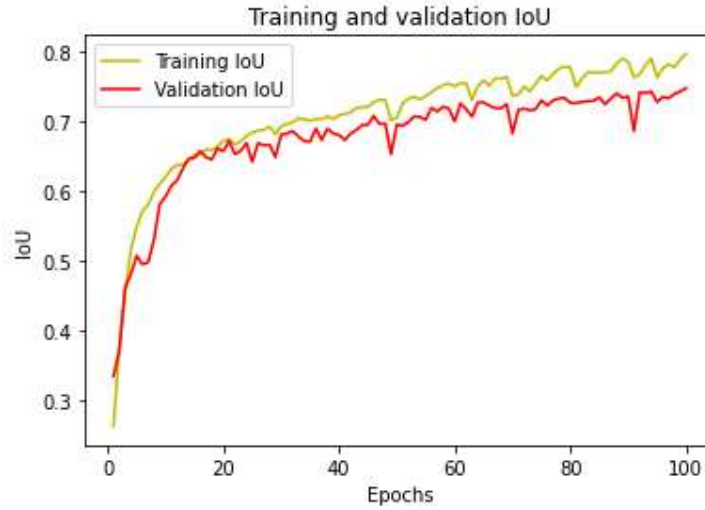
وجد أنه عند استخدام نسبة تسرب بقيمة معينة يقل تابع الخسارة في نهاية مرحلتي التدريب والاختبار (الشكل (9,10)) وبالتالي مؤشرات على عدم حدوث مشكلة *Overfitting* في ظل استخدام قاعدة بيانات صغيرة الحجم، بينما في حالة عدم استخدام نسبة تسرب (الشكل (8)) كما وجد تحسن ملحوظ بداية ومن ثم ارتفاع في قيمة تابع الخسارة عند اقتراب نهاية مرحلتي التدريب والاختبار.

2. تابع نسبة التقاطع:

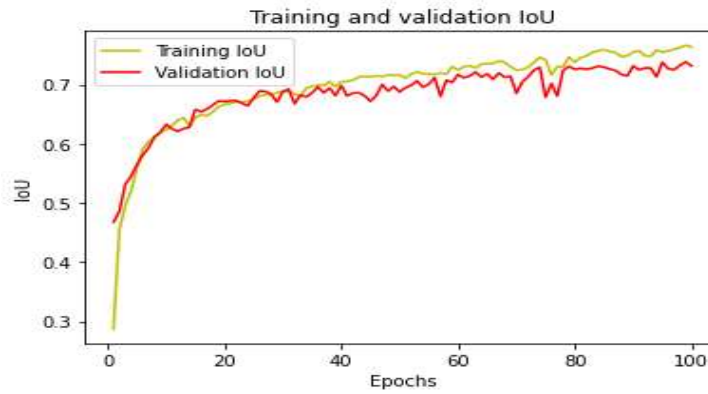
عند عدم استخدام نسب تسرب في مرحلة التدريب (وفق السيناريو الأول) ظهر في الشكل (11) انخفاض حاد في نسبة التقاطع في مرحلة الاختبار وذلك نتيجة صغر حجم قاعدة البيانات وبالتالي عدم قدرة النموذج على التعرف على صور جديدة خلال مرحلة الاختبار، وذلك على عكس حالة استخدام نسب تسرب (وفق السيناريو الثاني) مطبقة على خرج الطبقات الالتفافية حيث يظهر في الشكل (12) تحسن ملحوظ على تابع نسبة التقاطع في نهاية مرحلة الاختبار، أما في حالة تطبيق التسرب (وفق السيناريو الثالث) (الشكل (13)) فكان هناك تحسن خلال مرحلة الاختبار مع بداية انخفاض في نهاية المرحلة ذاتها. وبالتالي يمكن أن نقول أن أفضل النتائج لنسبة التقاطع حققت وفق السيناريو الثاني.



الشكل (11): تابع نسبة التقاطع للنموذج (100 epochs) بدون نسبة تسرب

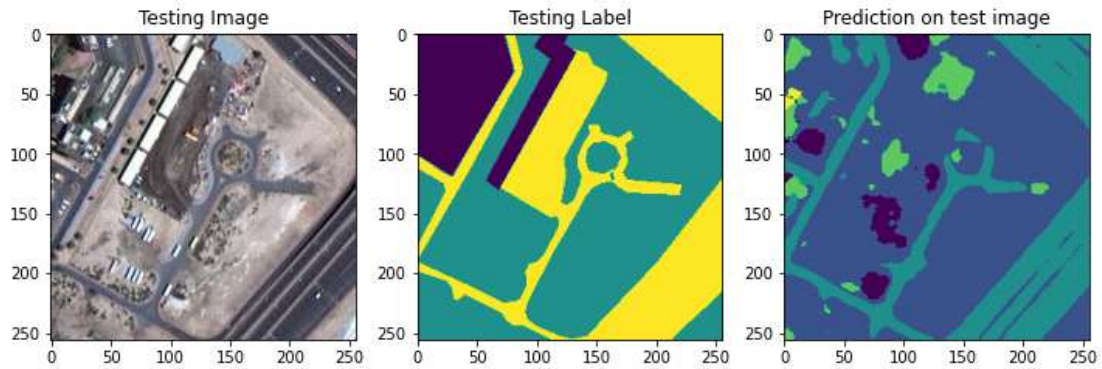


الشكل (12): تابع نسبة التقاطع للنموذج (100 epochs) مع نسبة تسرب



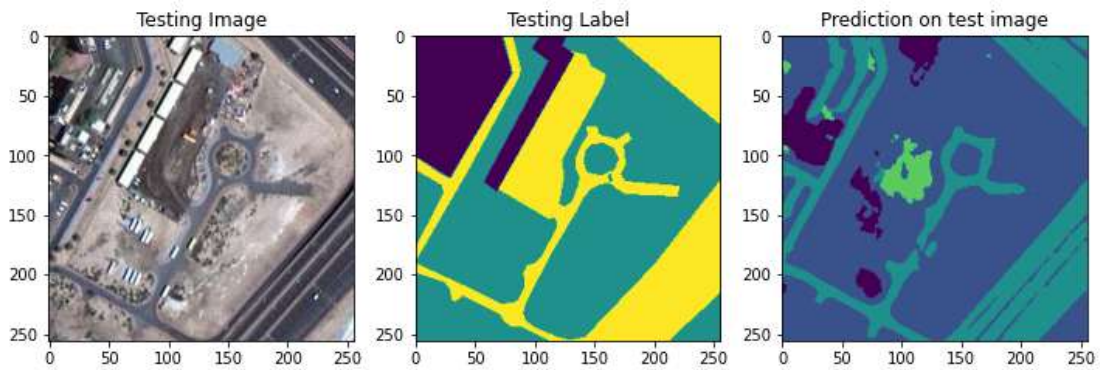
الشكل (13): تابع نسبة التقاطع للنموذج (100 epochs) مع تطبيق نسبة تسرب تصاعديا

أما كنتائج على دقة التقسيم الدلالي على الصور الفضائية مع وبدون استخدام نسب تسرب مختلفة، فلن تتأثر النتائج التي يمكن ملاحظتها بالعين المجردة.

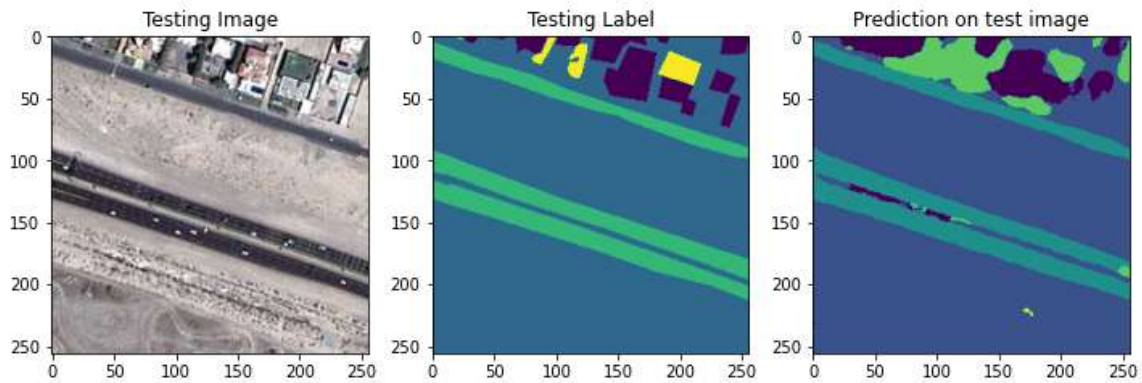


الشكل (14): صورة فضائية متوقعة مقسمة دلاليا بدون نسبة تسرب

ظهر في الشكل (13) والشكل (14) نفس الصورة مقسمة دلاليًا باستخدام نسب تسرب مختلفة، هناك تحسن بسيط باكتشاف بعض التفاصيل في الشكل (14) لم تكن موجودة في الشكل (13).

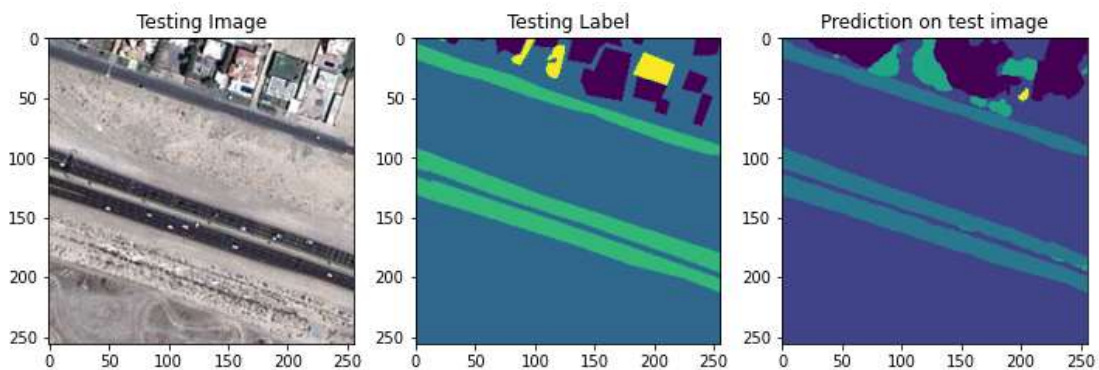


الشكل (15): صورة فضائية متوقعة مقسمة دلاليًا مع نسبة تسرب



الشكل (16): صورة فضائية متوقعة مقسمة دلاليًا بدون نسبة تسرب

ولكن في الشكل (15) والشكل (16) لم يظهر اختلاف في البكسلات التي تم تصنيفها دلاليًا باختلاف نسبة التسرب المطبقة.



الشكل (17): صورة فضائية متوقعة مقسمة دلاليًا مع نسبة تسرب

الاستنتاجات والتوصيات:

تم تدريب النموذج المقترح وفق ثلاث سيناريوهات، السيناريو الأول من دون تطبيق نسبة تسرب، السيناريو الثاني مع تطبيق نسبة تسرب 0.2 في كل الطبقات و0.3 في طبقة العنق، أما السيناريو الثالث بتطبيق نسبة تسرب ابتداءً من الطبقة الأولى 0.1 وتزداد بمقدار 0.1 عند كل طبقة. تم استخدام قاعدة بيانات مكونة من 72 صورة بأحجام مختلفة مما شكّل تحدي إضافي من ناحية معالجة الصور لزيادة عددها ولكي تكون بنفس الحجم حيث تم تقسيم البيانات إلى 80% تدريب و20% اختبار.

يظهر في الجدول (3) أنه في حال استخدام السيناريو الأول كان وقت التدريب هو الأقل بين السيناريوهات الثلاثة المقترحة حيث استغرق 20 ساعة فقط، وأعطى نسبة تقاطع في مرحلة التدريب 88% بينما في مرحلة الاختبار 70% كما هو موضح في الشكل (11)، في السيناريو الثاني ازداد وقت التدريب إلى 32 ساعة وكان تابع نسبة التقاطع في مرحلة التدريب 79% بينما في مرحلة الاختبار 74% وهذا ما يوضحه الشكل (12)، أما في السيناريو الثالث فقد ازداد وقت التدريب إلى 35 ساعة تدريب وأعطى تابع نسبة التقاطع قيمة 76% في مرحلة التدريب وقيمة 73% في مرحلة الاختبار كما هو موضح في الشكل (13).

وبما أن المشكلة الأساسية هي مشكلة فرط التخصص والتي تظهر في مرحلة الاختبار (عند ادخال بيانات جديدة لم يتم تدريب النموذج عليها) وبمقارنة قيم تابع نسبة التقاطع في مرحلة الاختبار في الجدول (3) بين السيناريوهات الثلاثة فيظهر أن السيناريو الثاني يعطي أفضل القيم مع أحداث بعض التقدم القليل نسبياً في جودة التقسيم الدلالي لبعض الصور كما في الشكل (14) والشكل (15)، وبعضها الآخر لم يظهر فيها تحسن ولكن لم يؤثر سلباً على دقة التقسيم الدلالي كما في الشكل (16) والشكل (17).

وفقاً للنتائج السابقة يمكن القول أنّ اعتماد السيناريو الثاني هو الأكثر تناسباً مع قاعدة البيانات المعتمدة في البحث لتجنب مشكلة فرط التخصص وتحقيق نسبة تقاطع أفضل كما يمكن متابعة البحث بشكل أكثر شمولاً للطبقات الأكثر تأثراً بتعديل نسبة التسرب.

References:

- [1] Saifi, M, Y; Singla, J; Ita, N." Deep Learning based Framework for Semantic Segmentation of Satellite Images". Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE Xplore Part Number: CFP20K25-ART; ISBN:978-1-7281-4889-2
- [2] BUDUMA, N.; LACASCIO, N. "Fundamentals of Deep Learning". June 2017.
- [3] Zhang, A.; Lipton, Z.; Li, Mu. & Smola, A., "Dive into Deep Learning". Sep 18, 2020.
- [4] Muruganandham, S." Semantic Segmentation of Satellite Images using Deep Learning". Czech Technical University in Prague & Luleå University of Technolog. August,2016.
- [5] WANG, Y; ZHANG, D; DAI, G." CLASSIFICATION OF HIGH RESOLUTION SATELLITE IMAGES USING IMPROVED U-NET". Int. J. Appl. Math. Comput. Sci., 2020, Vol. 30, No. 3.
- [6] Ronneberger, O; Fischer, P; Brox, T." U-Net: Convolutional Networks for Biomedical Image Segmentation". May 2015. volume 9351 of LNCS, pages 234–241.
- [7] Oktay,O; Schlemper,J; Le Folgoc,L; Lee,M; Heinrich,M; Misawa,K; Mori,K; McDonagh,S; Hammerla,N;Kainz,B;Glocker,B;Rueckert,D."Attention U-Net: Learning Where to Look for the Pancreas".

- [8] Seo,H; Huang,C; Bassenne,M; Xiao,R; Xing,L."Modified U-Net (mU-Net) with Incorporation of Object-Dependent High Level Features for Improved Liver and Liver-Tumor Segmentation in CT Images". IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 00, NO. 00.
- [9] Badrinarayanan, V; Kendall, A; Cipolla, R; Senior Member, IEEE." SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". Oct 2016.
- [10] Long, J; Shelhamer, E; Darrell, T." Fully Convolutional Networks for Semantic Segmentation".2015.
- [11] Geirhos, R; Janssen, D; Schütt, H; Rauber, J; Bethge, M; Wichmann, F." Comparing deep neural networks against humans: object recognition when the signal gets weaker". Dec 2018.
- [12] Khryashchev, V; Ivanovsky, L; Pavlov, V; Ostrovskaya, A; and Rubtsov, A. "Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation". THE 23RD CONFERENCE OF FRUCT ASSOCIATION.
- [13] Koga, Y; Miyazaki, H; and Shibasaki, R." A Method for Vehicle Detection in High-Resolution Satellite Images that Uses a Region-Based Object Detector and Unsupervised Domain Adaptation". 9 February 2020.
- [14] Rakhlin, A; Davydow, A; Nikolenko, S." Land Cover Classification from Satellite Imagery with U-Net and Lovasz-Softmax Loss". CVPR Workshop Paper.2019.
- [15] Parente, L; Taquary, E; Silva, A; Souza, C; and Ferreira, L." Next Generation Mapping: Combining Deep Learning, Cloud Computing, and Big Remote Sensing Data". 3 December 2019.
- [16] PHAN, T; NGUYEN, T; HOANG, T; NGUYEN, Q; and JO, J." Multi-Scale Bushfire Detection from Multi-Modal Streams of Remote Sensing Data". Vietnam National Foundation for Science and Technology Development (NAFOSTED). December 22, 2020.
- [17] Ba, R; Chen, C; Yuan, J; Song, W; and Lo, S." SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention". 18 July 2019.
- [18] Garg, L; Shukla, P; Singh, S; Bajpai, V; and Yadav, U." Land Use Land Cover Classification from Satellite Imagery using mUnet: A Modified Unet Architecture ". International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019), pages 359-365.
- [19]Srivastava,N;Hinton,G;Krizhevsky,A;Sutskever,I; Salakhutdinov,R. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting".2014
- [20] [Semantic segmentation of aerial imagery | Kaggle](#)
- [21] Yi Lin,T; Goyal,P; Girshick,R; He,K; Dollar,P."Focal Loss for Dense Object Detection".2017.
- [22] Srivastava,N; Hinton,G; Krizhevsky,A; Sutskever,I; Salakhutdinov,R." Dropout: A Simple Way to Prevent Neural Networks from Overfitting".2014.