

Table of Contents

10	قائمة المصطلحات العلمية:
14	الفصل 1: المقدمة
14	1.1. خلفية البحث و أهميته:
16	1.2. الدراسات المرجعية:
16	التعلم الآلي:
20	التعلم العميق:
22	التجزئة الدلالية:
25	تقدير العمق:
27	كشف النار قبل استخدام الذكاء الصناعي:
28	تحديد العمق دون استخدام الذكاء الصناعي:
30	خلاصة:
31	أهداف البحث:
31	قيود البحث:
31	مخطط الأطروحة:
34	الفصل 2: الإدراك البصري عند البشر
35	2.1. كيف يدرك البشر الصور مقابل كيفية فعل الآلات
36	سرعة معالجة المعلومات البصرية
37	التطورات في إدراك الصور الآلي
40	صيغ وخوارزميات ضغط الصور
43	الفصل 3: الشبكات العصبونية الالتفافية
44	معايير اختيار الشبكة العصبية الالتفافية:
46	الطبقة الالتفافية Convolutional Layer:
47	وحدة الانحدار الخطي المصححة (ReLU):
49	تابع التنفيع Sigmoid:
50	طبقة الاقتراع (Pooling Layer):
51	وصلات التخطي Skip Connections:
56	بنية الشبكة الشهيرة U-Net:
59	الفصل 4: مجموعات البيانات Datasets
61	2. مجموعة البيانات FLAME:
62	3. مجموعة بيانات KITTI:

63	مجموعة بيانات NYU Depth	4.
65	مجموعة بيانات DOIDE	5.
68	الفصل 5: التجارب والنتائج	
68	التجارب:	
69	المشاكل والحلول	
69	المعالجة المسبقة للبيانات (Data Pre-Processing):	
69	اختيار النموذج (Model Selection):	
69	ضبط الهايبربارامترات (Hyperparameters Tuning):	
70	إفراط التعلم (Overfitting):	
71	النتائج	
73	سليبيات النظام:	
74	الفصل 6: الخاتمة	
74	ملخص:	
74	الأعمال المستقبلية	

قائمة المصطلحات العلمية:

Fire Detection	اكتشاف النار
Depth Estimation	تقدير العمق
Machine Learning	التعلم الآلي
Deep Learning	التعلم العميق
Classification Models	نماذج التصنيف
Segmentation Masks	أقنعة التجزئة
Supervised Learning	التعلم بإشراف
Unsupervised Learning	التعلم من دون إشراف
Reinforcement Learning	التعلم المعزز
Linear Regression	الانحدار الخطي
Logistic Regression	الانحدار اللوجستي
Decision Trees	أشجار القرار
Random Forests	الأشجار العشوائية
Support Vector Machines	آلة شعاع الدعم
Data-Driven Decisions	القرارات المقادة بالبيانات
Stochastic Gradient Descent	الانحدار التدرجي العشوائي

17	Figure 1: خوارزميات التعلم الآلي
18	Figure 2: تطبيقات تعلم الآلة
23	Figure 3: معمارية شبكات الالتفاف الكلي (FCN)
23	Figure 4: بنية شبكة U-Net
24	Figure 5: بنية شبكة SegNet
24	Figure 6: بنية شبكة Mask RCNN
25	Figure 7: بنية شبكة Eigen et al
26	Figure 8: بنية نموذج [18]
26	Figure 9: نموذج الشبكة المقترحة Yang et al
27	Figure 10: حساس لاكتشاف النار
28	Figure 11: حساس دخان
29	Figure 12: رسم توضيحي لطريقة عمل حساسات Time-of-Flight
29	Figure 13: رسم توضيحي لطريقة حساب العمق من الصور الثنائية
30	Figure 14: مثال من قاعدة بيانات NYU والتي تم استنباط خرائط العمق فيها من حساسات LIDAR
34	Figure 15: Human Visual System
36	Figure 16: Order of visual signal processing steps
37	Figure 17: MobileNet Architecture
38	Figure 18: Generative Adversarial Networks Architecture
40	Figure 19: Variational Auto Encoders Architecture
41	Figure 20: JPEG Algorithm
41	Figure 21: PNG Compression Algorithm
43	Figure 22: Basic Convolutional Neural Network for Handwritten Digit Classification
46	Figure 23: Convolution Operation
47	Figure 24: ReLU Activation Function Graph
49	Figure 25: Sigmoid Activation Function
51	Figure 26: Average Pooling
52	Figure 27: Basic Identity Skip Connection
53	Figure 28: CNN Training Setup
56	Figure 29: Results Enhancement Techniques
57	Figure 30: U-Net Architecture
59	Figure 31: How data fits into Machine Learning Projects
60	Figure 32: Images Sample from FSID dataset (left is the original RGB image and right is the corresponding segmentation mask)
61	Figure 33: Frame samples of the normal spectrum palette
61	Figure 34: Frame samples of thermal images including Fusion, WhiteHot, and GreenHot
62	Figure 35: Segmentation Sample
62	Figure 36: Image Sample from KITTI Dataset
64	Figure 37: NYU Depth Dataset Sample
65	Figure 38: Sample of Images from DOIDE dataset
66	Figure 39: Data pre-processing pipeline
71	Figure 40: Table of Results
72	Figure 41: Prediction results (top: original Images, center: fire masks, bottom: depth masks)

يعد الكشف عن الحرائق وتحديد مواقعها أولوية حاسمة للسلامة العامة. لتلبية هذه الحاجة، قمنا بتطوير نموذجين يعتمدان على بنية شبكة UNet لاكتشاف الحرائق وتحديد مواقعها. يستخدم النموذج الأول بنية وحدة فك التشفير ويتم تدريبه على مجموعة بيانات Fire Segmentation Image Dataset و FLAME لتجزئة الحرائق. تم تدريب النموذج الثاني على تقدير العمق باستخدام مجموعات بيانات KITTI و NYU Depth v2 و DIODE. تظهر نتائجنا أن النموذج الأول يحقق متوسط دقة 80٪ في مهمة الكشف عن الحريق، بينما حقق النموذج الثاني معدل خطأ تربيعي متوسط قدره 0.09 في مهمة موقع الحريق. توضح هذه النتائج إمكانات تقنيات الذكاء الاصطناعي في تعزيز قدرات الكشف عن الحرائق وتحديد المواقع.

1.1. خلفية البحث و أهميته:

يتناول هذا الفرع الفرصة لتوضيح خلفية البحث وأهميتها في مجال اكتشاف الحرائق، وتعريف القارئ بمشكلة الحرائق وأثرها السلبي على المجتمع والاقتصاد. يتم تسليط الضوء على الأبحاث الحالية والتقنيات المستخدمة في اكتشاف الحرائق (Fire Detection)، بما في ذلك تقنيات التعلم الآلي (Machine Learning) والتعلم العميق (Deep Learning) ونماذج التصنيف (Classification Models) وتقدير العمق (Depth Estimation). كما يتم استعراض الأبحاث السابقة التي تتناول موضوع اكتشاف الحرائق قبل تطور التقنيات الحديثة، وأبحاث تقدير العمق قبل تطبيق التقنيات الحالية. تم تحديد أهداف البحث والتحديات التي ستواجهها الدراسة. وأخيراً، سيتم تقديم خطة البحث المتعلقة بتطوير نظام اكتشاف الحرائق الذي يقدر موقع الحريق وعمقه باستخدام تقنيات التعلم الآلي والتعلم العميق ونماذج التصنيف وتقدير العمق.

يسعى الإنسان دائماً إلى المضي قدماً نحو مستقبل متطور، وذلك بالاعتماد على التقنيات الحديثة التي تمكنه من تحقيق أهدافه بطريقة أفضل وأكثر فعالية. ومن بين هذه التقنيات، نجد أنظمة الحماية والمراقبة التي تعتمد على التعلم الآلي والتعلم العميق في اكتشاف الحرائق وتقدير موقعها وعمقها، مما يسهم في تحرير الإنسان من المهام الروتينية البسيطة والمملة وتركز الجهود على المهام الأكثر أهمية والتي تتعلق بسلامة الإنسان والحفاظ على بيئة العمل والمنزل الجيدة. وبالتالي، يتم تعزيز الكفاءة والإنتاجية في العمل، وتحسين جودة الحياة بشكل عام، وتوفير الوقت والجهد والمال الذي يستغرقه الإنسان في القيام بهذه المهام يدوياً. ومن هنا، يمكن القول إن أنظمة الحماية والمراقبة الذكية ستلعب دوراً مهماً في المستقبل، وستحقق تقدماً كبيراً في مجال الحماية والأمان والتحكم في البيئات المختلفة.

تطبيقات نظام كشف الحرائق وتقدير موقعها وعمقها متعددة، ومن بينها:

- **المباني السكنية والتجارية:** تعد المباني السكنية والتجارية من أهم الأماكن التي يجب حمايتها من الحرائق، حيث إن وجود حريق يمكن أن يؤدي إلى خسائر مادية كبيرة، وأحياناً يمكن أن يؤدي إلى فقدان الأرواح. ومن خلال نظام كشف الحرائق وتقدير موقعها وعمقها، يمكن تحديد موقع الحريق بدقة والتحكم فيه بشكل فعال قبل انتشاره، وبالتالي يتم حماية المبنى ومن فيه من الأضرار الناجمة عن الحرائق.

- **المصانع والمستودعات:** تعد المصانع والمستودعات من الأماكن التي يجب حمايتها من الحرائق بشكل خاص، حيث تحتوي على العديد من المواد الخطرة وربما المواد القابلة للاشتعال، ومن هنا، يتم وضع نظام كشف الحرائق وتقدير موقعها وعمقها للتحكم في أي حريق قد يحدث بشكل فعال وفي وقت قريب، وبالتالي يمكن تقليل الخسائر الناجمة عن الحرائق والحفاظ على المواد المخزنة بشكل آمن.
- **الغابات والمناطق الطبيعية:** تعد الغابات والمناطق الطبيعية من المواقع التي تتعرض للحرائق بشكل متكرر، الأمر الذي يؤثر سلباً على الحياة الحيوانية والنباتية، وأحياناً يمكن أن يؤدي إلى فقدان الممتلكات الخاصة والعامة. وبوجود نظام كشف الحرائق وتقدير موقعها وعمقها، يتم الكشف عن الحرائق في وقت مبكر والتحكم فيها بشكل أكثر فعالية عن طريق إجراء تكامل بين نظام كشف حرائق ونظام إطفاء.
- **المركبات وأنظمة النقل:** يمكن أن تكون الحرائق في المركبات وأنظمة النقل خطيرة بشكل خاص وتؤدي إلى فقدان الأرواح والأضرار الكبيرة. ومن خلال نظام كشف الحرائق الذي يستخدم أقنعة التجزئة (Segmentation Masks) وتقدير العمق، يمكن اكتشاف الحرائق بسرعة وتحديد موقعها داخل المركبة، مما يتيح اتخاذ إجراء فعال وسريع لمنع حدوث مزيد من الأضرار أو الإصابات. ويمكن أن يكون ذلك مهماً بشكل خاص لأنظمة النقل العامة، مثل القطارات أو الحافلات، حيث يمكن أن يكون عدد كبير من الأشخاص في خطر.
- **الفعاليات الخارجية والتجمعات العامة:** يمكن أن تكون الفعاليات الخارجية والتجمعات العامة، مثل الحفلات والمهرجانات والألعاب الرياضية، عرضة بشكل خاص للحرائق بسبب الأعداد الكبيرة من الناس والمصادر المحتملة للاشتعال العرضية أو العمدية. ويمكن نشر نظام كشف الحرائق الذي يستخدم أقنعة التجزئة وتقدير العمق في هذه الفعاليات لاكتشاف وتحديد أي حرائق قد تحدث بسرعة، مما يتيح الاستجابة السريعة لمنع المزيد من الأضرار أو الإصابات. ويمكن أن يساعد ذلك على ضمان سلامة الحضور وحماية أي مبانٍ أو هياكل قريبة من الأضرار.

1.2. الدراسات المرجعية:

التعلم الآلي:

تقنية التعلم الآلي أو (Machine Learning) تمثل تقدماً تكنولوجياً مهماً في السنوات الأخيرة. إنها تقنية فرعية من الذكاء الاصطناعي تتيح للحواسيب التعلم والتحسين بشكل ذاتي دون الحاجة إلى برمجة صريحة. لقد حوّلت تقنية التعلم الآلي العديد من الصناعات بما في ذلك الرعاية الصحية والمالية والتجزئة والتصنيع من خلال تمكين المؤسسات من استخلاص رؤى ذات مغزى من البيانات. تقنية التعلم الآلي هي فرع من فروع الذكاء الاصطناعي، تتيح للحواسيب التعلم والتحسين بشكل ذاتي عبر الخبرة. وهي عملية تدريب الحاسوب على التعلم من الأنماط الموجودة في البيانات، ثم استخدام هذه الأنماط لإجراء تنبؤات أو اتخاذ قرارات. يمكن لخوارزميات التعلم الآلي التعلم من مجموعات البيانات وتحسين دقتها مع مزيد من البيانات، والتي تسمى بيانات التدريب. كلما كانت لدى الخوارزمية المزيد من بيانات التدريب، كلما تحسنت -نظرياً على الأقل- قدرتها على التعلم والتنبؤ. [1]

أنواع تقنية التعلم الآلي:

توجد ثلاثة أنواع رئيسية لتقنية التعلم الآلي: التعلم بإشراف (Supervised Learning)، والتعلم من دون إشراف (Unsupervised Learning)، والتعلم المعزز (Reinforcement Learning). [1]

- ينطوي التعلم بإشراف (Supervised Learning) على استخدام بيانات تدريب مسمّاة (Labeled Training Data) لتدريب النماذج الآلية، حيث يتم توفير الإجابة الصحيحة للنموذج خلال عملية التدريب، وبناءً على ذلك يقوم النموذج بتوليد الإجابة المناسبة عندما يتم تقديمه للبيانات الجديدة. ويتم استخدام التعلم بإشراف في تصنيف الصور، والتعرف على الأشخاص والأغراض، وتحديد نوع النص الذي يتم تقديمه، والترجمة التلقائية، والتعرف على الكلام. [2]

- أما التعلم من غير إشراف (Unsupervised Learning) فيقوم بتعلم النموذج بدون الحصول على بيانات مسمّاة، وفي هذه الحالة يقوم النموذج بتحديد الأنماط والتكراريات في البيانات بدون الحصول على أي معلومات حول الإجابة الصحيحة. ويتم استخدام التعلم الغير مشرف في تجميع البيانات، وتحليل المحتوى، وكشف الأخطاء، وتحليل الصور والنصوص. [3]

- أما التعلم المعزز (Reinforcement Learning) فيقوم بتدريب النموذج من خلال تزويده بإشارات مكافآت (Reward Signals) أو عقوبات (Penalties) عندما يتم تنفيذ الإجراءات المختلفة في

بيئة محددة. ويقوم النموذج بتعلم كيفية اتخاذ القرارات المناسبة لتحقيق هدف محدد من خلال تفاعلاته مع البيئة المحيطة به. ويتم استخدام التعلم المعزز في الألعاب، والروبوتات، والتحكم في العمليات الصناعية.[4]

الخوارزميات في التعلم الآلي: [1]

تعتبر خوارزميات التعلم الآلي هي عناصر بناء النماذج المستخدمة في التعلم الآلي. وهناك العديد من خوارزميات التعلم الآلي، يتم تصميم كل منها لحل مشكلة محددة. وبعض أشهر خوارزميات التعلم الآلي هي:

1. الانحدار الخطي (Linear Regression) [5]: يستخدم لتوقع متغير مستمر. وهو خوارزمية بسيطة تفترض علاقة خطية بين المتغيرات المدخلة والمتغير الناتج.

2. الانحدار اللوجستي (Logistic Regression) [5]: يستخدم لتوقع النتائج الثنائية. وهو خوارزمية بسيطة تستخدم وظيفة السيغمويد لربط المتغيرات المدخلة بالمتغير الناتج.

3. أشجار القرار (Decision Trees) [6]: خوارزمية تعتمد على الأشجار وتستخدم لحل مشاكل الانحدار والتصنيف. حيث يتم تقسيم البيانات إلى مجموعات أصغر بناءً على مجموعة من القواعد.

4. الغابات العشوائية (Random Forest) [6]: يتكون هذا النوع من عدة أشجار قرارية ويستخدم لحل مشاكل الانحدار والتصنيف. حيث يتم دمج توقعات العديد من الأشجار القرارية لتحسين دقة النموذج.

5. آلة شعاع الدعم (Support Vector Machine) [7]: يستخدم لحل مشاكل الانحدار والتصنيف. حيث يتم العثور على أفضل فراغ بين البيانات يفصل نقاط البيانات إلى فئات مختلفة.

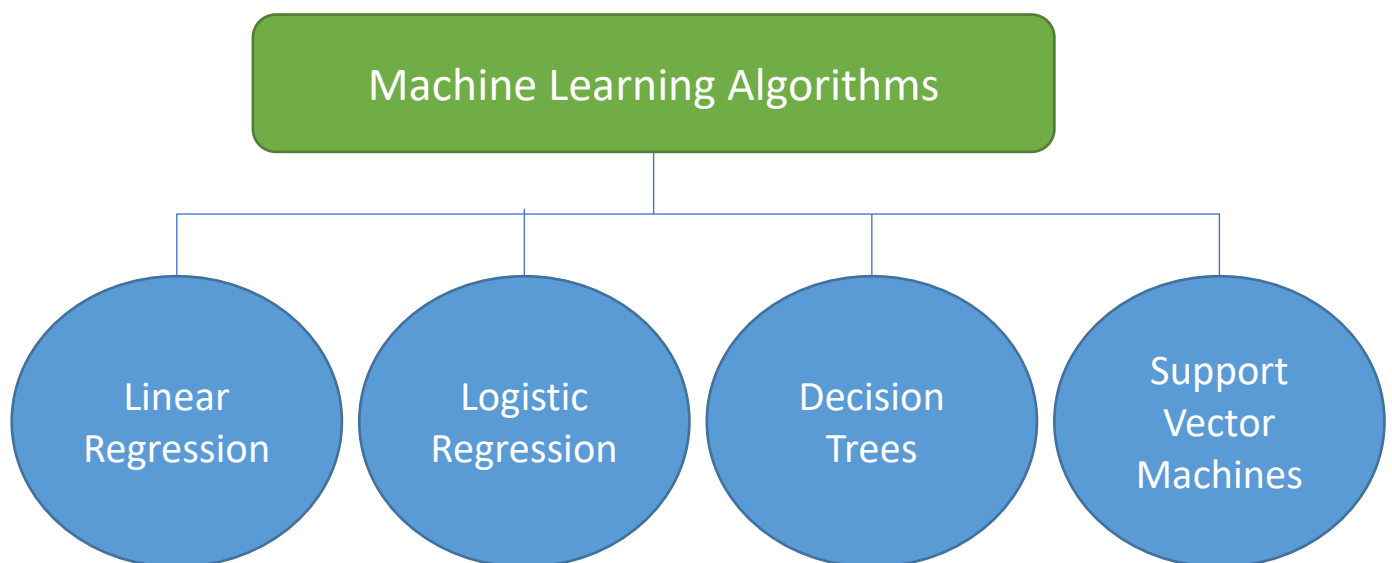


Figure 1: خوارزميات التعلم الآلي

تطبيقات التعلم الآلي:

يحتوي التعلم الآلي على العديد من التطبيقات في مختلف الصناعات، بما في ذلك الرعاية الصحية والتمويل والتجزئة والتصنيع. وبعض أشهر تطبيقات التعلم الآلي هي: [1]

1. الصيانة التنبؤية (Predictive Maintenance): يمكن استخدام التعلم الآلي للتنبؤ بموعد فشل آلة ما. ويمكن أن يساعد ذلك المؤسسات على جدولة الصيانة قبل أن تتعطل الآلة.
2. الكشف عن الاحتيال (Fraud Detection): يمكن استخدام التعلم الآلي للكشف عن المعاملات الاحتيالية. ويمكن أن يحلّل الأنماط في البيانات لتحديد المعاملات المشبوهة.
3. التسويق المخصص (Customised Marketing): يمكن استخدام التعلم الآلي لتحليل بيانات العملاء وإنشاء حملات تسويقية مخصصة. ويمكن أن يساعد ذلك المؤسسات على زيادة مشاركة العملاء والولاء.
4. التشخيص الطبي (Medical Diagnostics): يمكن استخدام التعلم الآلي لتشخيص الحالات الطبية. ويمكن أن يحلّل بيانات المرضى ويقدم توصيات للمهنيين الصحيين.
5. السيارات الذاتية القيادة (Self-Driving Cars): يمكن استخدام التعلم الآلي لتطوير السيارات الذاتية. ويمكن أن يحلّل بيانات الاستشعار لاتخاذ قرارات بشأن التسارع والفرملة والتوجيه.

Machine Learning Applications

Predictive Maintenance

Fraud Detection

Customised Marketing

Medical Diagnostics

Self-Driving Cars

Figure 2: تطبيقات تعلم الآلة

تأثير التعلم الآلي:

لدي التعلم الآلي تأثير كبير على مختلف الصناعات. فقد سمح بتوليد رؤى من كميات كبيرة من البيانات واتخاذ القرارات التي تعتمد على البيانات. وفيما يلي بعض الطرق التي تأثر بها التعلم الآلي على مختلف الصناعات:

1. **الرعاية الصحية:** غير التعلم الآلي بشكل جذري قطاع الرعاية الصحية عن طريق تمكين الأطباء من تشخيص الحالات الطبية بشكل أدق وتطوير خطط العلاج الشخصية. كما ساعد في تحسين نتائج المرضى وتخفيض تكاليف الرعاية الصحية.

2. **الأموال:** غير التعلم الآلي بشكل جذري أيضاً قطاع الأموال والاستثمار عن طريق تمكين المؤسسات من كشف المعاملات الاحتيالية وتوقع اتجاهات السوق وتطوير محافظ الاستثمار الشخصية.

3. **التجزئة:** ساعد التعلم الآلي التجار التجزئة في تحسين مشاركة العملاء والولاء من خلال توفير توصيات وعروض شخصية. كما ساعد في تحسين إدارة المخزون وعمليات سلسلة التوريد.

4. **التصنيع:** غير التعلم الآلي قطاع التصنيع عن طريق تمكين الشركات من تحسين جودة المنتج وتخفيض تكاليف الإنتاج وتحسين عمليات سلسلة التوريد. كما ساعد في تحسين سلامة العمال عن طريق تحديد المخاطر المحتملة في عملية الإنتاج.

5. **النقل:** ساعد التعلم الآلي على تطوير السيارات ذاتية القيادة، مما يمكن أن يحسن أمن الطرقات ويساعد في تخفيف الاختناقات المرورية. كما قد ساعد في تحسين العمليات اللوجستية عن طريق توقع حجم الطلب وتحسين طرق التوصيل.

التحديات في التعلم الآلي:

على الرغم من فوائده العديدة، يواجه التعلم الآلي العديد من التحديات. من أكبرها، عدم الشفافية في نماذج التعلم الآلي. يمكن أن تكون نماذج التعلم الآلي معقدة، وقد يكون من الصعب فهم كيفية اتخاذها للقرارات. يمكن أن يجعل هذا النقص في الشفافية من الصعب التعرف على الأخطاء في النموذج وتصحيحها.[1]

كما يمكن أن تكون نماذج التعلم الآلي متحيزة إذا تم تدريبها على مجموعة بيانات محدودة. يمكن أن يؤدي هذا التحيز إلى توقعات غير دقيقة.

وأخيراً، يواجه التعلم الآلي أيضاً تحديات فيما يتعلق بالخصوصية والأمان. يمكن أن تكون نماذج التعلم الآلي عرضة للهجمات، ويمكن أيضاً استخدامها لانتهاك حقوق الخصوصية الفردية.

التعلم العميق:

التعلم العميق هو فرع من فروع التعلم الآلي الذي غير مجال الذكاء الاصطناعي بشكل جذري. في العقد الماضي، قاد التعلم العميق إلى إنجازات غير مسبوقة في مجالات مثل رؤية الحاسوب، معالجة اللغة الطبيعية، التعرف على الكلام، والروبوتات. التعلم العميق هو أداة قوية لحل المشاكل المعقدة التي كان يعتقد سابقاً أنها مستحيلة الحل باستخدام تقنيات التعلم الآلي التقليدية. [8]

يمكن تتبع أصول التعلم العميق إلى الأربعينات من القرن الماضي، عندما اقترح العالمان (Warren McCulloch and Walter Pitts) أول شبكة عصبية [9]. والفكرة وراء الشبكة العصبية هي محاكاة سلوك الدماغ البشري الذي يتألف من خلايا عصبية متصلة يمكنها معالجة ونقل المعلومات. ومع ذلك، فإن النهج الخاص بالشبكة العصبية سقط في النسيان في السبعينات بسبب نقص الطاقة الحاسوبية والبيانات.

في الثمانينات، تم تطوير خوارزمية الانتشار الخلفي (Backward Propagation) [10]، والتي سمحت للشبكات العصبية بالتعلم من البيانات عن طريق ضبط أوزانها والانحرافات الخاصة بها. وهذا أدى إلى تجديد الاهتمام بالشبكات العصبية، ولكن كان أداؤها محدوداً لا يزال بسبب البنية السطحية ومجموعات البيانات الصغيرة.

المفاهيم الأساسية:

يعتمد التعلم العميق على فكرة الشبكات العصبية الاصطناعية التي تتألف من عقد متصلة، أو خلايا عصبية، تقوم بمعالجة ونقل المعلومات. تتألف الشبكة العصبية من طبقة إدخال وطبقة أو أكثر خفية وطبقة خرج. يستلم كل عقد إشارات الإدخال من خلايا أخرى وينتج إخراجاً عن طريق تطبيق وظيفة تنشيط غير خطية. تُعد الأوزان والانحرافات الخاصة بالعقد مُعدلة خلال عملية التدريب لتقليل الخطأ بين الخرج المتوقع والخرج الحقيقي.

يتم عادة إجراء عملية التدريب باستخدام نوع من الانحدار التدرجي العشوائي (Stochastic Gradient Descent)، والذي ينطوي على تحديث الأوزان والانحرافات بشكل تكراري بناءً على قيمة مشتق تابع الكلفة بالنسبة للمعلمات. يتم استخدام خوارزمية الانتشار الخلفي لحساب التدرج بكفاءة عن طريق نقل الخطأ بشكل عكسي خلال الشبكة.

التطبيقات:

تم تطبيق التعلم العميق بنجاح على مجموعة واسعة من التطبيقات، بما في ذلك: [8]

- **رؤية الحاسوب:** أتاح التعلم العميق تقدماً كبيراً في التعرف على الصور والفيديو، والكشف عن الكائنات، والتقسيم. ومن بين التطبيقات المهمة: المركبات ذاتية القيادة، والتعرف على الوجوه، والصور الطبية.
- **معالجة اللغة الطبيعية:** أتاح التعلم العميق تقدماً كبيراً في الترجمة الآلية، وتلخيص النصوص، وتحليل المشاعر. ومن بين التطبيقات المهمة: المساعدات الافتراضيين، وروبوتات المحادثة، ونماذج اللغة.
- **التعرف على الكلام:** أتاح التعلم العميق تقدماً كبيراً في التعرف على الكلام وتحليله، مما أدى إلى تطوير مساعدين صوتيين ونظم التحويل النصي للكلام وغيرها.
- **الروبوتيات:** أتاح التعلم العميق تقدماً كبيراً في إدراك الروبوتات، والتحكم فيها، والتخطيط لها، مما أدى إلى تطوير روبوتات ذاتية للتصنيع، والخدمات اللوجستية، والخدمات الأخرى.

التحديات:

على الرغم من نجاحاته، تواجه الشبكات العصبية العميقة العديد من التحديات، ومنها:

- **التحيز في البيانات:** تتأثر خوارزميات التعلم العميق بالتحيزات في بيانات التدريب، مما يؤدي إلى نتائج غير عادلة أو تمييزية. لحل هذا التحدي، يتطلب الأمر اختياراً دقيقاً للبيانات ومعالجتها، بالإضافة إلى التدخلات الخوارزمية والسياسية.
- **صعوبة التفسير:** تُعتبر النماذج العصبية العميقة غالباً "صناديق سوداء"، مما يجعل من الصعب فهم كيفية اتخاذها للقرارات. وتُعتبر التفسيرية أمراً هاماً لضمان الثقة والشفافية والمساءلة في أنظمة الذكاء الاصطناعي.
- **هجمات الخصوم (Adversarial Attacks):** تتعرض النماذج العصبية العميقة للهجمات الخصومية التي تنطوي على إجراء تغييرات صغيرة على بيانات الإدخال لإيهام النموذج بأن الإدخال يتعلق بشيء آخر. يتطلب الأمر العمل على تصميم نماذج عصبية عميقة قوية ومناهضة للهجمات الخصومية، في الآونة الأخيرة، نلاحظ هذا النوع من الهجمات على نماذج اللغات الضخمة مثل ChatGPT، وذلك لتحريض هذه النماذج على توجيه إساءة إلى فئات معينة من البشر، أو بشكل عام، التلاعب بالدخل لإجبار هذه النماذج على إنتاج نصوص مسيئة.

التجزئة الدلالية.

هي مهمة أساسية في رؤية الحاسوب والتي تتضمن تقسيم صورة إلى مناطق مختلفة ذات معنى دلالي. وهي تقنية رئيسية لتطبيقات مختلفة مثل كشف الكائنات، وتجزئة الصور، والقيادة الذاتية. في هذه المقالة، سنناقش بالتفصيل مفهوم التجزئة الدلالية وتقنياتها المختلفة وتطبيقاتها في رؤية الحاسوب. [11]

التجزئة الدلالية هي مهمة في الرؤيا الحاسوبية تتضمن تسمية كل بكسل في الصورة بتصنيف دلالي مقابل. بمعنى آخر، فإنه ينطوي على تسمية بكسل واحد في الصورة بتصنيف دلالي بناءً على معناه الدلالي. الهدف من التجزئة الدلالية هو تحديد مواقع الكائنات في الصورة عن طريق تقسيمها إلى عدة مناطق، حيث يتوافق كل منطقة مع كائن مختلف.

أهمية التجزئة الدلالية:

التجزئة الدلالية هي مهمة أساسية في الرؤيا الحاسوبية، ولها عدة تطبيقات في مجال تحليل الصور والفيديو. وبعض التطبيقات الأكثر شيوعاً للتجزئة الدلالية تتضمن ما يلي:

- **كشف الكائنات:** يمكن استخدام التجزئة الدلالية للكشف عن الكائنات في الصورة، من خلال تعيين تصنيف فريد لكل بكسل في الصورة.
- **القيادة الذاتية:** التجزئة الدلالية هي مهمة حاسمة في تطوير أنظمة القيادة الذاتية. حيث تسمح بتحديد وتتبع الكائنات في الوقت الحقيقي، مثل المشاة والمركبات وعلامات الطريق.
- **الصور الطبية:** يتم استخدام التجزئة الدلالية في الصور الطبية لتحديد مواقع هياكل تشريحية مختلفة في الجسم، مثل الأعضاء والأنسجة والعظام.
- **الروبوتيات:** يمكن استخدام التجزئة الدلالية في الروبوتيات لتحديد مواقع الكائنات في المشهد، وهو مهم للمهام مثل الإمساك والتلاعب.

تقنيات التجزئة الدلالية

هناك العديد من التقنيات المختلفة المستخدمة في التجزئة الدلالية، ولكل منها مزاياها وعيوبها. وبعض أشهر التقنيات تشمل:

Fully Convolutional Networks (CNNs): هي بنية شبكة عصبية تم تقديمها في عام 2015 للتجزئة الدلالية. وتستبدل طبقات الاتصال الكاملة (Fully Connected Layers) بطبقات التفاضلية أحادية الفلاتر مما يمكن من التدريب من البداية إلى النهاية للتجزئة الدلالية. [12]

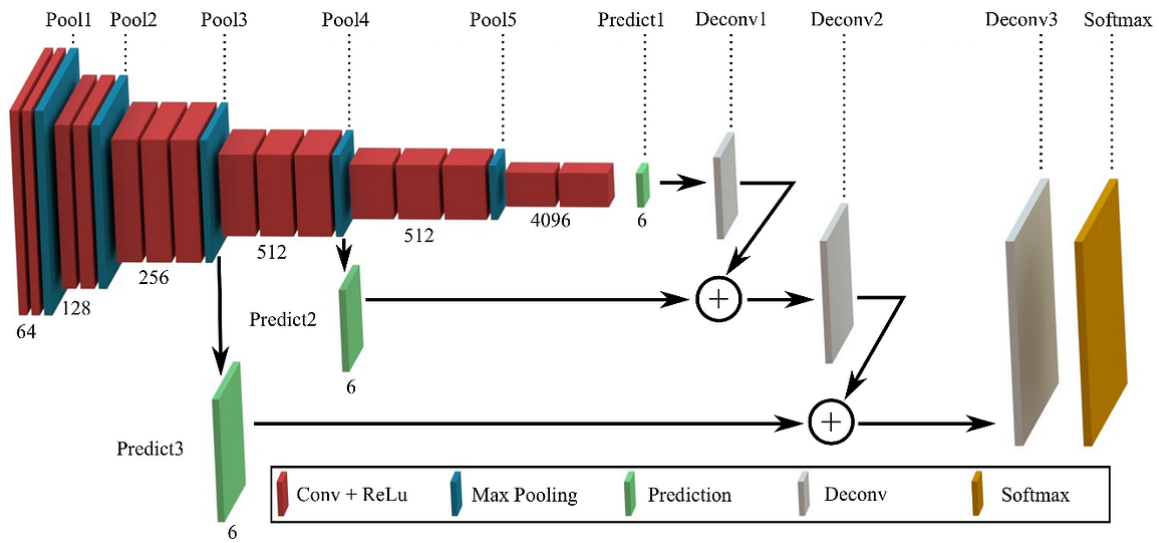


Figure 3 معمارية شبكات الالتفاف الكلي (FCN)

U-Net: هي بنية شبكة عصبية تم تقديمها في عام 2015 لتجزئة الصور الطبية. وتعتمد على بنية مرز-مفكك ترميز، حيث تقوم الطبقة المشفرة بتعلم تمثيل تسلسلي هرمي للصورة، بينما تقوم طبقة المفكك بتوليد خريطة التجزئة. [13]

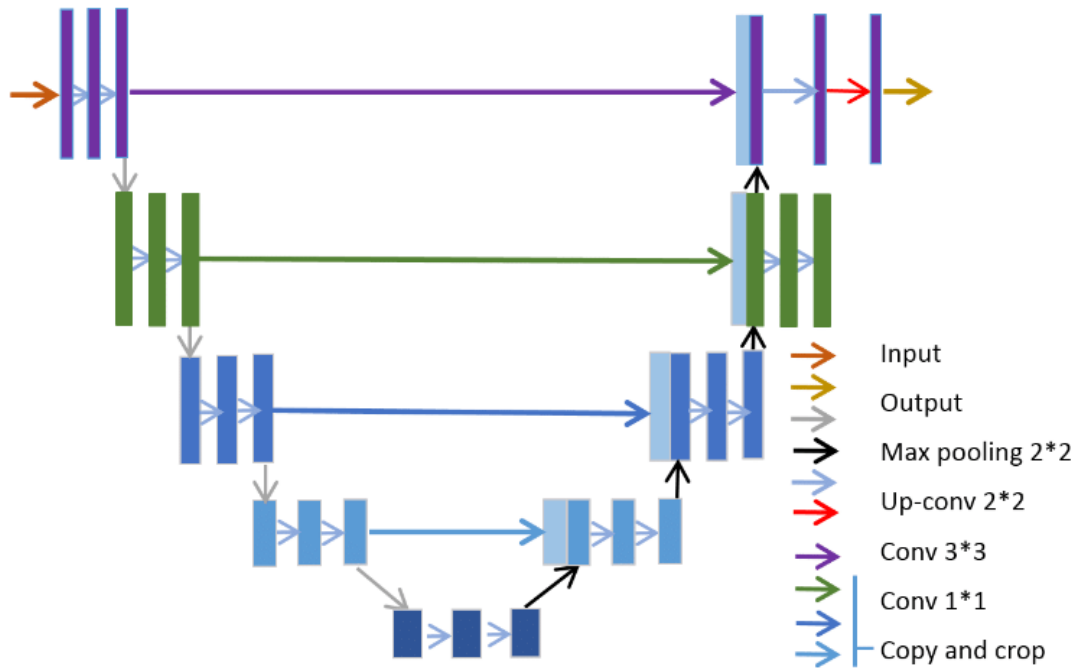


Figure 4 بنية شبكة U-Net

SegNet: هي معمارية شبكة عصبونية عميقة للتجزئة الدلالية تم تقديمها في عام 2016 لتحقيق التقسيمات الدلالية. تستخدم عمارة الترميز والتفكيك (كما في الشبكة المذكورة سابقاً) حيث يتعلم الترميز التمثيل الهرمي للصورة ويولد المفكك خريطة التقسيم.[14]

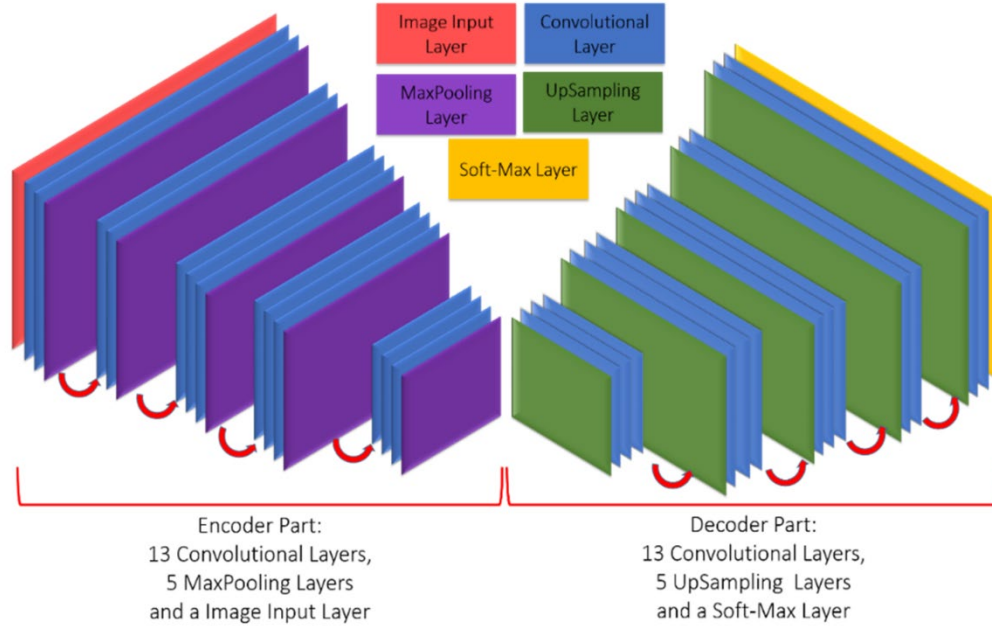


Figure 5 بنية شبكة SegNet

Mask RCNN: هي معمارية تعلم عميق تم تقديمها في عام 2018 لمهام تحديد الأغراض (Object Detection) وتحديد نمط الحواف (Edge Patterns). تعتمد على معمارية (Faster R-CNN) وتضيف فرعاً لتوقع قناع التشریح.[15]

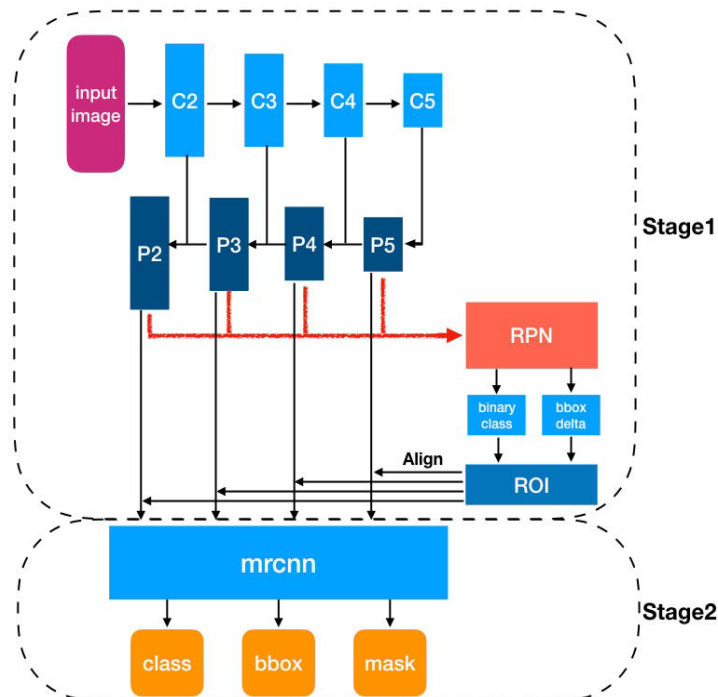


Figure 6 بنية شبكة Mask RCNN

تقدير العمق:

تقدير العمق هو مهمة أساسية في الرؤية الحاسوبية تتضمن تقدير مسافة الأجسام عن الكاميرا أو الحساس. تتضمن هذه المهمة العديد من التطبيقات في مجالات مثل الروبوتات والملاحة الذاتية والواقع الافتراضي والواقع المعزز. في السنوات الأخيرة، ظهرت التعلم العميق كتقنية قوية لتقدير العمق بسبب قدرتها على تعلم التمثيلات المعقدة من كميات كبيرة من البيانات. في هذه المقالة، سنناقش مختلف تقنيات التعلم العميق المعتمدة على تقدير العمق وتطبيقاتها.

تقدير العمق باستخدام صورة واحدة Monocular Depth Estimation:

تقدير العمق باستخدام صورة واحدة فقط مهمة صعبة حيث تفقد معلومات العمق في الصورة الثنائية الأبعاد. ومع ذلك، أظهرت تقنيات التعلم العميق نتائج مبشرة في هذه المهمة. واحدة من أقدم التقنيات المعتمدة على التعلم العميق لتقدير العمق باستخدام صورة واحدة هي العمل [16] حيث اقترحوا شبكة عميقة متعددة المقاييس تقوم بتوقع العمق من صورة واحدة. وأظهروا أن طريقتهم تفوقت على طرق الخصائص المصممة يدوياً التقليدية (Hand-Engineered Features).

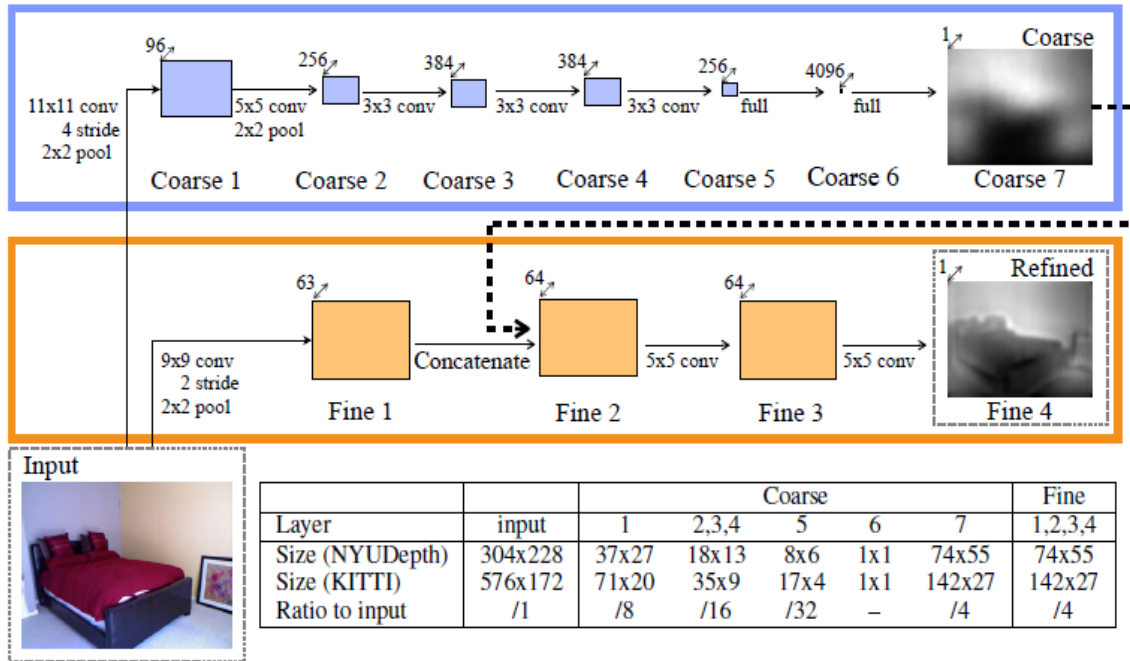


Figure 7 بنية شبكة Eigen et al

منذ ذلك الحين، تم اقتراح عدة تقنيات تعتمد على التعلم العميق لتقدير العمق من خلال صورة واحدة. اقترح Zhou وآخرون [17] شبكة عميقة تستخدم نهجاً متعدد المقاييس وطبقة تصحيح العمق الجديدة لتحسين دقة تقدير العمق.

تقدير عمق الصور المزدوجة:

تقدير عمق الإستريو هو مهمة تقدير العمق باستخدام زوج من صور (Stereo). تعتبر هذه المهمة أسهل نسبياً مقارنة بتقدير العمق من خلال صورة واحدة حيث تكون معلومات العمق ضمناً متوفرة في الصور المزدوجة (Stereo). ومع ذلك، لا يزال تقدير عمق الصور المزدوجة يشكل تحديات عدة، مثل التغطيات والانحرافات والضوضاء. تم استخدام تقنيات التعلم العميق للتعامل مع هذه التحديات وتحسين دقة تقدير عمق الإستريو.

واحدة من التقنيات التي تعتمد على التعلم العميق لتقدير عمق الصور المزدوجة العمل [18] الذين اقترحوا شبكة عميقة تتعلم لمطابقة البكسلات بين صور الإستريو وتقدير خريطة الانحراف. أظهرت أن طريقتهم تفوقت على الطرق التقليدية التي تعتمد على ملامح يدوية.

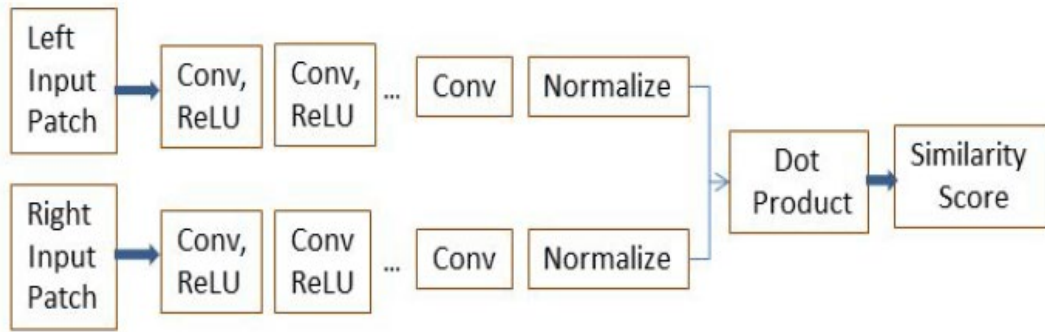


figure 8 بنية نموذج [18].

منذ ذلك الحين، تم اقتراح عدة تقنيات تعتمد على التعلم العميق لتقدير عمق الإستريو. اقترح Yang وآخرون (2018) شبكة عميقة تستخدم وحدة تجميع هرمية مكانية وآلية انتباه موجهة لتحسين دقة تقدير عمق الإستريو. اقترح Luo وآخرون (2019) شبكة عميقة تستخدم طبقة تحويل تحويل آفي محلية لتنظيم خريطة الانحراف وتحسين دقتها. [19]

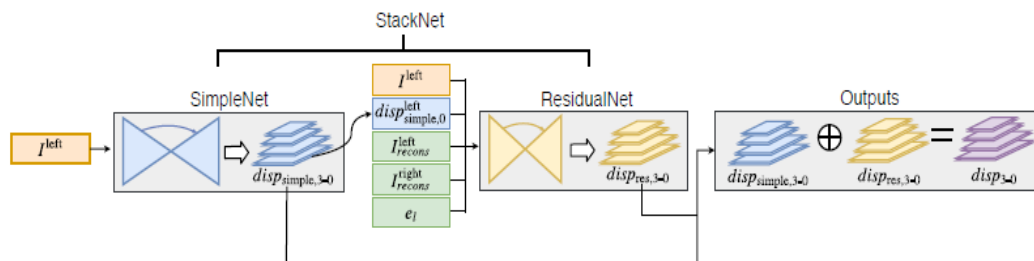


figure 9 نموذج الشبكة المقترحة Yang et al.

تقدير العمق هي مهمة أساسية في مجال الرؤيا الحاسوبية، وتمتلك تطبيقات عديدة في مختلف المجالات. ظهر التعلم العميق كتقنية قوية لتقدير العمق، وتم اقتراح العديد من التقنيات القائمة على التعلم العميق لتقدير العمق بالاستناد على صورة واحدة وصورتين. أظهرت هذه التقنيات نتائج واعدة وقد تحسنت دقة تقدير العمق. مستقبل تقدير العمق يتمثل في تطوير تقنيات قائمة على التعلم العميق أكثر دقة وكفاءة يمكن أن تتعامل مع تحديات السيناريوهات الحقيقية.

كشف النار قبل استخدام الذكاء الصناعي:

على الرغم من أن تقنيات التعلم الآلي والتعلم العميق هي تقنيات شائعة لكشف الحرائق، إلا أنه يوجد طرق أخرى لكشف الحرائق دون استخدام هذه التقنيات. في هذا المقال، سنستكشف بعض الطرق الممكنة لكشف الحرائق بدون استخدام التعلم الآلي والتعلم العميق. [20]

أحد الطرق المتبعة لكشف الحريق هو استخدام الحساسات. يمكن لحساسات الحرارة كشف الارتفاع المفاجئ في درجة الحرارة، مما يشير إلى وجود حريق. تكشف حساسات الدخان عن وجود الدخان، وهو مؤشر على وجود حريق. يمكن تثبيت هذه الحساسات في المباني ودمجها مع أجهزة الإنذار وأنظمة الرشاشات. يمكن لحساسات اللهب كشف وجود لهب ويمكن استخدامها بالتزامن مع حساسات الحرارة والدخان لتوفير نظام شامل لكشف الحرائق.



figure 10 حساس لاكتشاف النار

بالإضافة إلى الحساسات، هناك تقنيات أخرى يمكن استخدامها لكشف الحرائق. على سبيل المثال، يمكن للحساسات الصوتية كشف صوت الحريق، مثل صوت تقطيع اللهب أو صوت الأشياء التي تترق. يمكن لحساسات الغاز كشف وجود الغازات القابلة للاشتعال، والتي يمكن أن تكون مؤشراً على وجود حريق.



figure 11 حساس دخان

يمكن استخدام هذه الحساسات في البيئات الصناعية أو المناطق التي يوجد فيها مخاوف من تسرب الغاز. من المهم الإشارة إلى أن هذه التقنيات يمكن أن تكون فعالة في كشف الحرائق، ولكنها ليست مضمونة بالكامل. يمكن أن تتم تفعيل إنذارات كاذبة بسبب عوامل مثل الرطوبة العالية أو البخار، وقد لا يكون بعض الحرائق قابلة للكشف باستخدام هذه التقنيات. لذلك، من المهم وضع نظام متعدد للكشف عن الحرائق واختباره وصيانته بانتظام لضمان أنه يعمل بشكل صحيح.

تحديد العمق دون استخدام الذكاء الصناعي:

في السنوات الأخيرة، كانت الطرق العميقة للتعليم والتعلم الآلي هي الأساليب السائدة المستخدمة لتقدير العمق. ومع ذلك، هناك تقنيات وأجهزة استشعار أخرى يمكن استخدامها لتقدير العمق دون الاعتماد على الذكاء الاصطناعي البرمجي. [21]

إحدى التقنيات التي يمكن استخدامها لتقدير العمق هي كاميرات الوقت-الرحلة (Time-of-Flight (ToF)). تصدر هذه الكاميرات نبضة ضوء وتقيس الوقت اللازم للضوء للعودة إلى الكاميرا بعد أن يرتد

عن طريق جسم. من خلال حساب الوقت الذهاب والإياب، يمكن تحديد المسافة بين الكاميرا والجسم. وغالباً ما تُستخدم كاميرات ToF لتطبيقات مثل التعرف على الإيماءات، والروبوتات، وتتبع الأجسام.

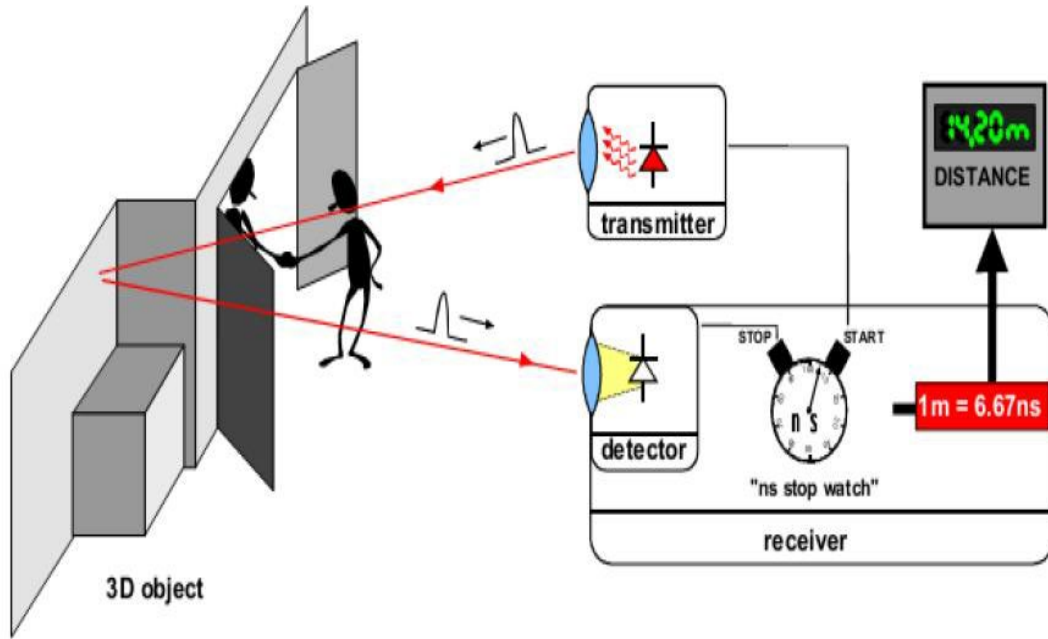


figure 12: رسم توضيحي لطريقة عمل حساسات Time-of-Flight

تقنية أخرى يمكن استخدامها لتقدير العمق هي رؤية الاستريو. تتضمن هذه التقنية استخدام كاميرتين لالتقاط نفس المنظر من زوايا مختلفة. من خلال تحليل الاختلاف بين الصور الملتقطة بواسطة الكاميرتين، يمكن تقدير المسافة بين الأجسام في المنظر. وقد استخدمت رؤية الاستريو في تطبيقات مثل إعادة الإنشاء ثلاثية الأبعاد وتخطيط العمق والروبوتات.

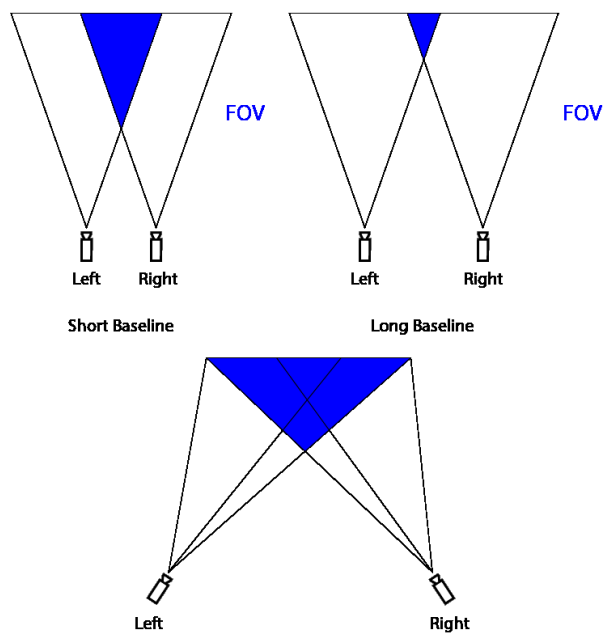


figure 13: رسم توضيحي لطريقة حساب العمق من الصور الثنائية

أخيراً، LIDAR (الكشف عن الضوء والتحديد بالرادار) هي تقنية يمكن استخدامها لتقدير العمق. تتضمن LIDAR إرسال شعاع ليزر وقياس الوقت الذي يستغرقه الليزر للعودة بعد أن يصطدم بجسم. من خلال حساب وقت الطيران، يمكن تحديد المسافة بين الكاميرا والجسم. تم استخدام LIDAR في تطبيقات مثل القيادة الذاتية ورسم الخرائط ثلاثية الأبعاد والروبوتات.

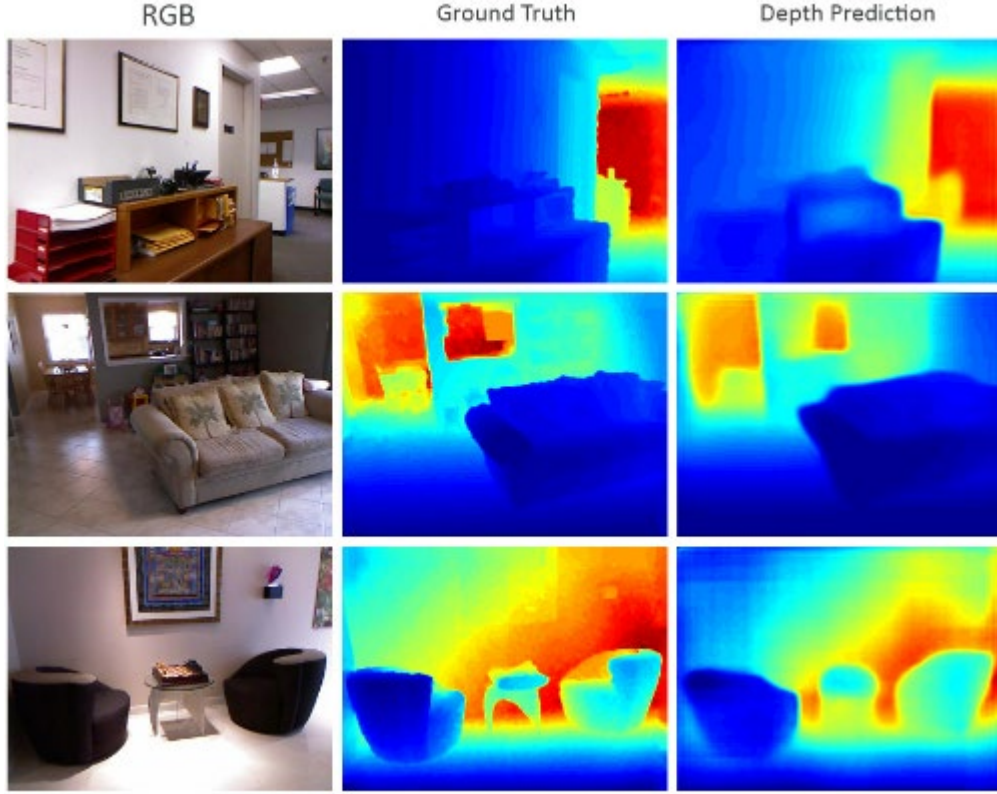


figure 14 مثال من قاعدة بيانات NYU والتي تم استنباط خرائط العمق فيها من حساسات LIDAR

خلاصة:

عملية اكتشاف النار هي عملية أكثر تعقيداً من عمليات تحديد الأغراض الأخرى Object Detection حيث أن النار ليست عنصراً ذو شكل أو حجم ثابتين وبالتالي تحديد النار في صورة هي مسألة يتم التعامل معها كمسألة تجزئة Segmentation بدلاً من أن تكون مسألة اكتشاف صندوق التحديد Bounding Box Detection، إضافة إلى ذلك، تحديد موقع النار بدقة هو أمر يزداد صعوبة عن الاعتماد على الصور فقط، حيث أن قواعد البيانات Datasets المتاحة لأغراض البحث العلمي لا تنطبق إلى مثل هذه المسائل فكان من الممكن إيجاد قاعدة بيانات للتعرف على النار وأخرى لمسألة تقدير العمق، ولكن لا يوجد قاعدة بيانات لتحديد عمق النار مما سيتطلب المزيد من المعالجة لبناء فهم رقمي ثلاثي الأبعاد للصورة ليتم تحليلها من قبل نماذج ذكاء اصطناعي أخرى.

أهداف البحث:

تتلخص أهداف البحث فيما يلي:

- استخدام خوارزميات التعلم العميق لكشف النار وتحديد الأجزاء من الصورة Segments التي تحتوي على نار.
- استخدام خوارزميات التعلم العميق لبناء خريطة عمق Depth Map ليتم استخدامها في تحديد بعد النار عن الكاميرا المستخدمة.

قيود البحث:

- **قيود العتاد:** حيث يتطلب تدريب نظام الذكاء الاصطناعي بالتعلم العميق معالجة كبيرة للغاية من قبل الحواسيب، وذلك بسبب العديد من العمليات الحسابية المعقدة التي يقوم بها الخوارزميات. وبالتالي، يحتاج تدريب هذا النظام إلى حواسيب ذات معالجات قوية وسرعات معالجة كبيرة.
- **قيود حجم قاعدة البيانات:** حيث يتطلب استخدام الشبكات العصبية العميقة قواعد بيانات ضخمة حتى تتدرب بشكل جيد وتعطي نتائج دقيقة. ولذلك، يعتبر إعداد قاعدة البيانات لتدريب هذا النظام عملية مرهقة وتستغرق وقتاً طويلاً، كما أن حجم القواعد الكبير يؤدي إلى زيادة زمن المعالجة.

مخطط الأطروحة:

الفصل الأول: المقدمة

الخلفية والأهمية البحثية: شرح أهمية تطوير نظام كشف الحرائق القائم على الذكاء الاصطناعي باستخدام تقنيات تقدير العمق في تحسين تدابير الأمان والحماية.

الدراسات المرجعية: تقديم لمحة عامة عن الأدبيات المتعلقة، بما في ذلك الأبحاث السابقة حول تعلم الآلة والتعلم العميق وتقنيات التشكيل الدلالي وتقنيات تقدير العمق.

الأهداف البحثية: تحديد بوضوح أهداف البحث وأهداف الدراسة.

القيود البحثية: مناقشة القيود والتحديات التي يواجهها البحث، مثل توافر البيانات ودقة الخوارزمية والموارد الحسابية.

خطة البحث: تحديد منهجية البحث والنهج لتحقيق أهداف البحث، بما في ذلك جمع البيانات والمعالجة الأولية واختيار النموذج والتقييم.

الفصل الثاني: التعرف البصري لدى الإنسان والآلة

المقدمة: مناقشة الاختلافات بين كيفية إدراك الإنسان للصور وكيفية تعرف الآلة على الصور، بما في ذلك العمليات الفسيولوجية والمعرفية المشاركة.

صيغ الصور وخوارزميات الضغط: تقديم لمحة عن صيغ الصور المختلفة وخوارزميات الضغط المستخدمة في معالجة الصور وتأثيرها على التعرف البصري.

الفصل 3: الشبكات العصبية التكرارية ومعمارياتها

المقدمة: تقدم نظرة عامة على الشبكات العصبية التكرارية المتزامنة (CNNs) وأهميتها في التعلم العميق لمهام معالجة الصور.

معايير اختيار الشبكة العصبية التكرارية المتزامنة: مناقشة المعايير المستخدمة لاختيار أفضل شبكة عصبية لنظام كشف الحرائق، بما في ذلك الدقة والكفاءة الحسابية وتعقيد النموذج.

المدخلات والطبقات المستخدمة في الشبكات العصبية التكرارية المتزامنة: شرح أنواع المدخلات والطبقات المختلفة المستخدمة في الشبكات العصبية التكرارية المتزامنة، بما في ذلك المدخلات الصورة والطبقة التكرارية ووحدة الانحدار الخطي المعكوسة (ReLU) ووظيفة التنشيط Sigmoid والطبقة التجميع وربط التخطي والطبقة التكرارية المتزامنة المعكوسة.

هندسة الشبكة: وصف الهندسة العامة للشبكة العصبية المستخدمة في نظام الكشف عن الحرائق وتبرير قرارات التصميم.

تدريب الشبكة: مناقشة عملية التدريب وضبط المعلمات المستخدمة لتحسين أداء الشبكة العصبية.

الفصل الرابع: مجموعات البيانات

ما هي مجموعة البيانات: شرح مفهوم مجموعة البيانات في التعلم الآلي وأهمية اختيار مجموعة بيانات تمثيلية للتدريب والتقييم.

مجموعة بيانات تفصيلية للتحليل الناري: شرح مجموعة البيانات المستخدمة للتحليل الناري، بما في ذلك جمع البيانات وتعليقها ومعالجتها مسبقاً.

مجموعة بيانات لتحديد العمق: شرح مجموعة البيانات المستخدمة لتحديد العمق، بما في ذلك جمع البيانات وتعليقها ومعالجتها مسبقاً.

مرحلة معالجة البيانات: مناقشة تقنيات معالجة البيانات المستخدمة لتطبيع وتعزيز مجموعات البيانات للتدريب والتقييم.

الفصل 5: التجارب والنتائج

التجارب: شرح إعداد التجربة وعملية اختبار نظام كشف الحريق باستخدام تكوينات مختلفة للشبكة العصبية الاصطناعية.

المشاكل والحلول: مناقشة التحديات والقيود التي واجهت خلال مرحلة الاختبار والحلول المتبعة للتغلب عليها.

النتائج: تقديم نتائج نظام كشف الحريق من حيث الدقة والحساسية والتخصصية والكفاءة الحسابية.

عيوب نظام الاهتمام: مناقشة القيود والعيوب لنظام كشف الحريق والمجالات المحتملة للتحسين.

الفصل 6: الأعمال المستقبلية والتوصيات

الملخص: خلاصة للنتائج والمساهمات الرئيسية للبحث وإعادة صياغة الأهداف والأهداف البحثية.

الأعمال المستقبلية: اقتراح الاتجاهات المحتملة للأبحاث المستقبلية في تحسين نظام اكتشاف الحرائق، مثل إدماج أنواع أخرى من الأجهزة الاستشعارية أو تحسين بنية CNN.

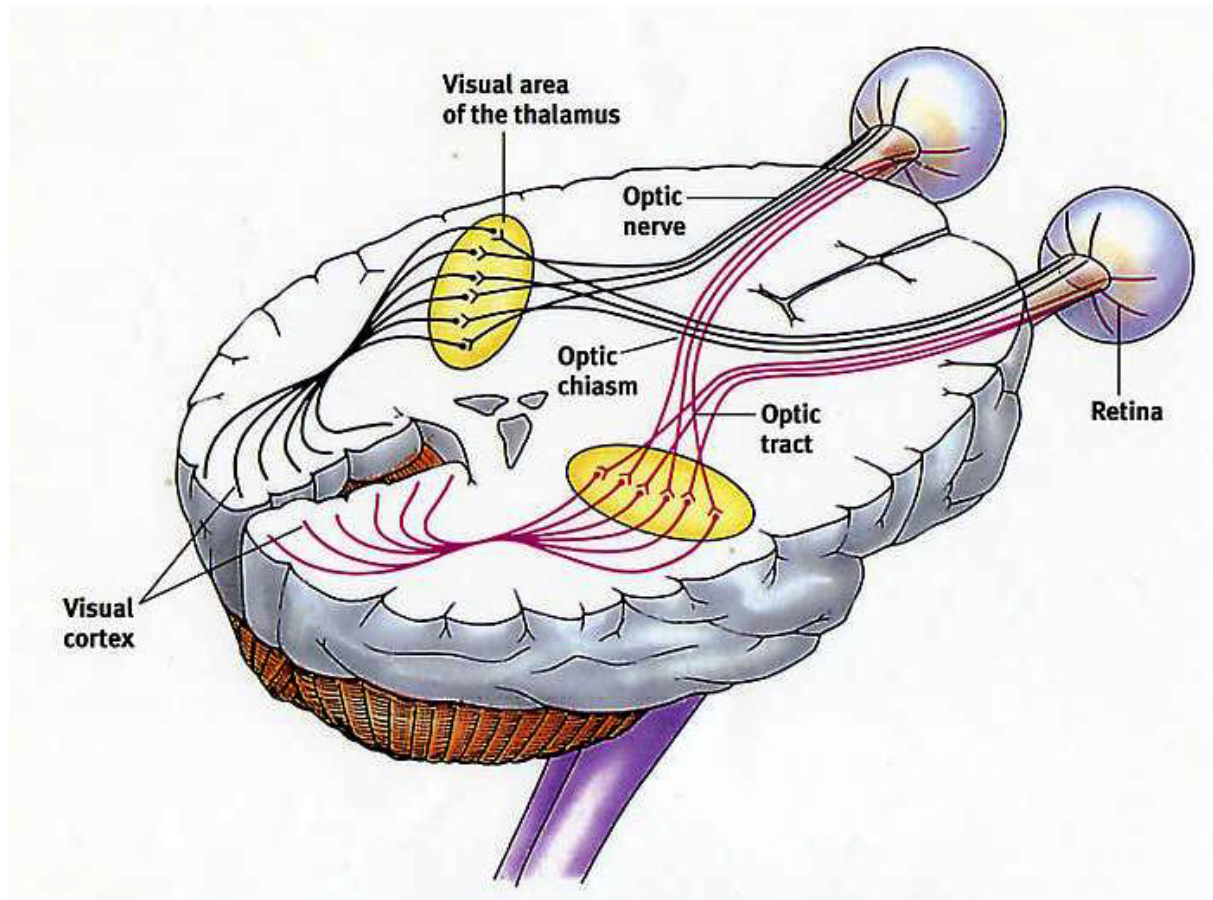
مجالات التحسين: تقديم توصيات لتحسين عملية جمع البيانات والتعليق، وتحسين تقنيات المعالجة المسبقة، وتحسين توجيه المعلومات.

الفصل 2: الإدراك البصري عند البشر

الإدراك البصري هو قدرة الدماغ البشري على تفسير وفهم المعلومات البصرية من البيئة. وهو عملية معقدة تشمل العينين والعصب البصري ومناطق مختلفة من الدماغ. إن الإدراك البصري البشري هو جانب أساسي في الحياة اليومية، حيث يتيح للأفراد تصفح محيطهم، وتعرف الأشياء والوجوه، وتفسير المعلومات البصرية بدقة.

دراسة الإدراك البصري لفتت منذ فترة طويلة اهتمام العلماء والباحثين في مختلف المجالات. في علوم الحاسوب وتعلم الآلة، فإن فهم الإدراك البصري البشري يعد حاسماً لتطوير خوارزميات ونماذج قادرة على معالجة وتفسير المعلومات البصرية بطريقة تشبه إلى حد كبير الإدراك البشري. في الروبوتيات، يعد الإدراك البصري أمراً أساسياً لتطوير الروبوتات التي يمكنها التنقل والتفاعل مع البيئة بطريقة مشابهة لكيفية تفاعل البشر.

التعرف البصري هو فرع من فروع الإدراك البصري يركز على قدرة البشر والآلات على التعرف على الأشياء والمشاهد في الصور والفيديوهات. يتضمن هذه العملية تحليل المعلومات البصرية لتحديد الأشياء وخصائصها وعلاقتها بالأشياء الأخرى في المشهد. يعد التعرف البصري أساسياً في العديد من التطبيقات مثل المركبات ذاتية القيادة والبحث في الصور والفيديو والتصوير الطبي.



Human Visual System15 figure

2.1. كيف يدرك البشر الصور مقابل كيفية فعل الآلات

يعتبر الإدراك البصري جزءاً أساسياً من تجربة الإنسان. فهو يسمح لنا بجمع ومعالجة المعلومات البصرية، وفهم العالم من حولنا، واتخاذ القرارات المستنيرة. ومع استمرار تقدم التكنولوجيا، تصبح الآلات أيضاً قادرة على فهم المعلومات البصرية بشكل متزايد. لقد كان إدراك الآلات للصور موضوعاً للبحث منذ عدة عقود، ولكنه اكتسب اهتماماً كبيراً في السنوات الأخيرة مع ظهور خوارزميات التعلم العميق. [22]

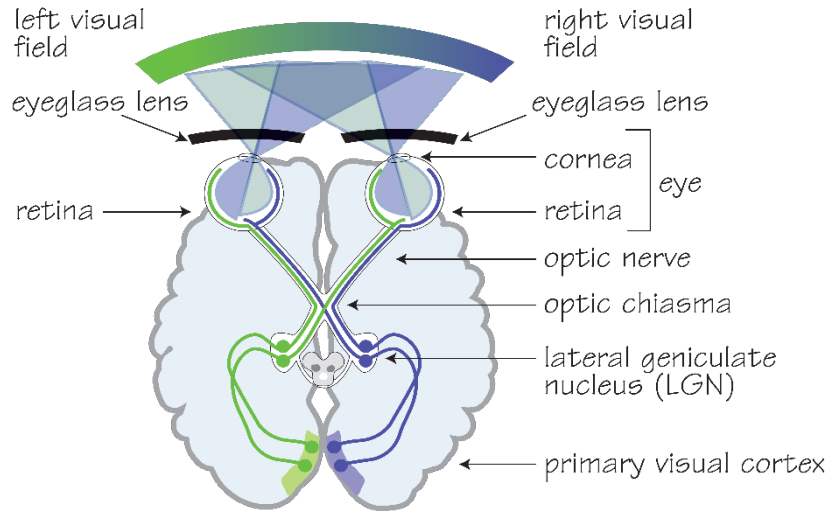
بينما هناك تشابهات بين إدراك الإنسان والآلات للصور، إلا أن هناك أيضاً اختلافات جوهرية. يستطيع البشر استيعاب الصور بطريقة بديهية وطبيعية، بينما تتطلب الآلات مجموعة من الخوارزميات والنماذج الرياضية لأداء نفس المهمة. فقد جاءت قدرة الإنسان على استيعاب الصور والتعرف على الأنماط نتيجة ملايين السنين من التطور، بينما يعد إدراك الآلات للصور مجالاً جديداً نسبياً لا يزال في مهده.

إدراك الإنسان للصور هو عملية معقدة ومتطورة تتطوي على مراحل متعددة. تلتقط العين البشرية المعلومات البصرية وترسلها إلى الدماغ، حيث يتم معالجتها وتفسيرها. ينشئ الدماغ بعد ذلك تمثيلاً للمشاهد البصري، والذي يعتمد على مزيج من المدخلات الحسية والمعرفة السابقة. هذه العملية من الإدراك البصري أساسية للإنسان للتنقل والتفاعل مع العالم من حوله.

المرحلة الأولى من الإدراك البصري تتطوي على العين. تتكون العين البشرية من عدة هياكل تعمل معاً لالتقاط الضوء وتركيزه. تعمل القرنية والعدسة معاً على ثني الضوء الوارد وتركيزه على الشبكية، وهي طبقة رقيقة من الأنسجة في الخلفية الداخلية للعين. تحتوي الشبكية على خلايا متخصصة تسمى الخلايا الضوئية، والتي تحول الضوء الوارد إلى إشارات كهربائية يمكن نقلها إلى الدماغ.

عندما يتم التقاط المعلومات البصرية بواسطة العين، يتم إرسالها إلى الدماغ عبر العصب البصري. يتم معالجة المعلومات البصرية أولاً في القشرة البصرية الأساسية، التي تقع في الفص الصدغي للدماغ. تتحمل القشرة البصرية الأساسية المسؤولية عن معالجة المعلومات البصرية الأساسية، مثل الحواف والخطوط والأشكال.

من القشرة البصرية الأساسية، يتم إرسال المعلومات البصرية إلى مناطق معالجة بصرية أعلى في الدماغ. تتولى هذه المناطق معالجة معلومات بصرية أكثر تعقيداً، مثل الوجوه والأشياء والمشاهد. كما يشارك هذه المناطق في العمليات مثل الانتباه والذاكرة واتخاذ القرارات، والتي تعد حاسمة لتفسير المعلومات البصرية في السياق.



Order of visual signal processing steps16 figure

واحدة من أهم قوى إدراك الإنسان للصور هي القدرة على التعرف على الأنماط والأشياء في المشهد. يمكن للإنسان بسرعة تحديد الأشياء وفهم العلاقات المكانية الخاصة بها بناءً على الإشارات البصرية مثل الشكل والحجم واللون والقوام. تعتمد هذه القدرة على قدرة الدماغ على التعرف والتصنيف على الأشياء بناءً على المعرفة والخبرة السابقة.

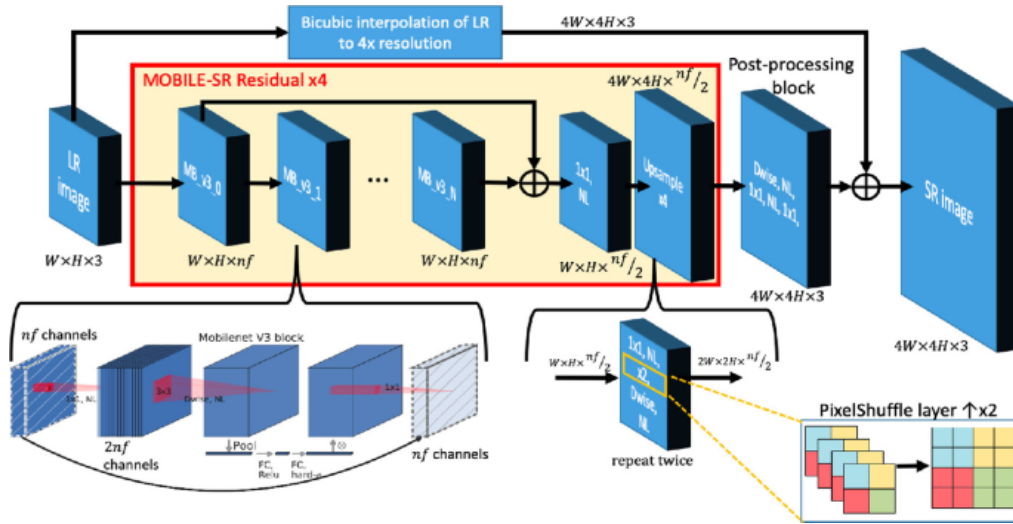
جانب آخر مهم للإدراك البصري البشري هو القدرة على تفسير المعلومات البصرية في سياقها. يستطيع البشر استخدام معرفتهم بالعالم لتفسير المشاهد البصرية بطريقة معنوية. على سبيل المثال، عندما يشاهد البشر صورة للشاطئ، يستطيعون استخدام معرفتهم بالعالم لتحديد الرمال والماء والسماء. تستند هذه القدرة على تكامل الدماغ للمدخلات الحسية مع المعرفة السابقة والخبرة.

سرعة معالجة المعلومات البصرية

أحد الاختلافات الرئيسية بين إدراك الإنسان والآلة للصور هو السرعة التي يتم بها معالجة المعلومات البصرية. الدماغ البشري قادر على معالجة المعلومات البصرية بسرعة كبيرة، مما يتيح لنا القدرة على التعرف بسرعة على الأشياء في المشهد وفهم علاقاتها المكانية.

بالمقابل، يمكن أن يكون إدراك الآلة للصور بطيئاً نسبياً، حيث تحتاج خوارزميات التعلم العميق إلى كميات كبيرة من الطاقة الحسابية لمعالجة المعلومات البصرية. ومع ذلك، فقد تم تحسين سرعة خوارزميات التعلم العميق بشكل كبير في الآونة الأخيرة بفضل التقدم في تقنية الأجهزة، مما يتيح معالجة الصور بشكل فوري في بعض التطبيقات.

بالإضافة إلى تحسينات الأجهزة، هناك جهود لتحسين خوارزميات التعلم العميق للحصول على معالجة أسرع. على سبيل المثال، طوّر بعض الباحثون معماريات أكثر كفاءة لـ شبكات العصب الاصطناعي المتسلسلة، مثل (U-Net و SegNet)، والتي تقلل من المتطلبات الحسابية للخوارزميات مع الحفاظ على الدقة. [13][14]



MobileNet Architecture17 figure

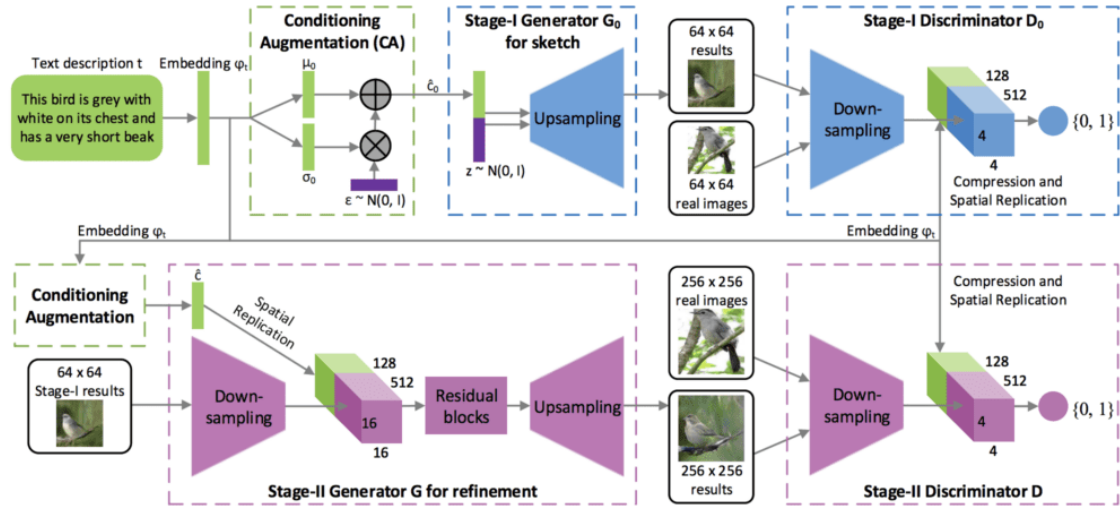
على الرغم من هذه التحسينات، ما زالت معالجة الصور بواسطة الآلة تتخلف عن الإدراك البصري للإنسان من حيث السرعة. ويمكن أن يكون هذا الأمر عاملاً محدداً في بعض التطبيقات، مثل المركبات الذاتية، حيث يكون معالجة الصور في الوقت الحقيقي أمراً حاسماً للعمل الآمن.

التطورات في إدراك الصور الآلي

في السنوات الأخيرة، شهدنا تطورات مهمة في إدراك الصور الآلي، وخاصة في مجال الرؤية الحاسوبية. فقد أدى تطوير خوارزميات التعلم العميق، والتي تعتمد على الشبكات العصبية الصناعية، إلى اختراقات في التعرف والتصنيف الصوري. تم تصميم خوارزميات التعلم العميق لمحاكاة هيكل ووظيفة الدماغ البشري، مما يمكن الآلة من التعلم من مجموعات بيانات كبيرة من الصور وتعرف على الأنماط والكائنات.

واحدة من أهم التطورات في تصور الصور من قبل الآلات هو تطوير الشبكات العصبية المركزة على التباينات (CNNs) وهي عبارة عن نوع من خوارزميات التعلم العميق متخصصة بشكل خاص في مهام التعرف على الصور. تم تصميم شبكات CNN للكشف عن الأنماط في صورة الإدخال عن طريق تطبيق مجموعة من المرشحات على الصورة. تكتشف كل مرشحة ميزة محددة، مثل حافة أو زاوية، ويتم دمج إخراج هذه المرشحات لإنشاء تمثيل للصورة يستخدم للتصنيف. [23]

تطور آخر في إدراك الآلة للصور هو تطوير الشبكات التوليدية التنافسية (Generative Adversarial Networks GANs) وهي نوع من خوارزميات التعلم العميق المستخدمة لتوليد الصور. تتكون GANs من شبكتين عصبيتين: شبكة المولد وشبكة المميز. يتم تدريب شبكة المولد لإنشاء صور جديدة مشابهة لمجموعة محددة من صور التدريب، بينما تتم تدريب شبكة المميز على التمييز بين الصور الحقيقية والصور المولدة. يتم تدريب الشبكتين بشكل متزامن، حيث يحاول شبكة المولد خداع شبكة المميز لتصنيف صورها المولدة على أنها حقيقية. []



Genrative Adversarial Networks Architecture18 figure

حققت التطورات في فهم الآلة للصور تحسناً كبيرة في مجموعة واسعة من التطبيقات، بدءاً من المركبات الذاتية إلى الصور الطبية. على سبيل المثال، يتم استخدام شبكات العصب العميق المتسلسلة في المركبات الذاتية لتحديد الكائنات في البيئة، مثل المشاة والمركبات الأخرى. وتستخدم شبكات مولد الصراع الإنتاجية في الصور الطبية لإنتاج صور اصطناعية للأعضاء، والتي يمكن استخدامها لتدريب خوارزميات تعلم الآلة على اكتشاف وتشخيص الأمراض.

مع ذلك، على الرغم من هذه التطورات، ما زالت هناك تحديات كبيرة في تحليل الصور من خلال الآلة. واحدة من هذه التحديات هي الحاجة إلى كميات كبيرة من البيانات المصنفة. تتطلب خوارزميات التعلم العميق كميات كبيرة من البيانات للتدريب بشكل فعال، وتصنيف هذه البيانات يمكن أن يستغرق وقتاً ويكلف الكثير من المال. ومن التحديات الأخرى هي الحاجة إلى موارد الحوسبة ذات الأداء العالي. تحتاج خوارزميات التعلم العميق إلى طاقة حاسوبية كبيرة للتدريب والتشغيل، والأمر يمكن أن يكون عائقاً أمام اعتمادها من قبل المؤسسات الصغيرة.

في الخلاصة، هناك تقدمات كبيرة في فهم الآلات للصور في السنوات الأخيرة، بفضل تطوير خوارزميات التعلم العميق والشبكات العصبية الاصطناعية. وقد أدت هذه التقدمات إلى اختراقات في التعرف على

الصور والتصنيف، وتستخدم في مجموعة واسعة من التطبيقات. ومع ذلك، هناك تحديات لا يزال يجب التغلب عليها، ولا سيما في مجال تصنيف البيانات وموارد الحوسبة ذات الأداء العالي.

2.2. كيفية تخزين الصور من قبل الإنسان مقابل الآلة

يتمتع الدماغ البشري بقدرة على معالجة وتخزين كميات هائلة من المعلومات البصرية، بينما تتمتع الآلات بسعة محدودة لتخزين ومعالجة البيانات. وعلاوة على ذلك، تختلف طريقة تخزين الصور من قبل الإنسان والآلة بشكل كبير.

يقوم الإنسان بتخزين المعلومات البصرية في ذاكرته الطويلة الأمد، والتي يمكن استرجاعها في وقت لاحق. ويتم تخزين المعلومات بشكل موزع عبر الدماغ، حيث تتحمل مناطق مختلفة من الدماغ مسؤولية معالجة جوانب مختلفة من الرؤية، مثل اللون والشكل والحركة. وتتميز عملية تخزين واسترجاع المعلومات البصرية لدى الإنسان بالمرونة العالية، مما يسمح بالتعرف على الأشياء حتى تحت ظروف الرؤية المختلفة. [23]

على النقيض، تخزن الآلات الصور على شكل مصفوفات من الأرقام، حيث يمثل كل رقم قيمة اللون لبكسل محدد في الصورة. هذا النهج فعال للغاية في تخزين ومعالجة كميات كبيرة من البيانات. ومع ذلك، فهو غير مرن ولا يسمح بنفس المستوى من التعرف تحت ظروف الرؤية المختلفة.

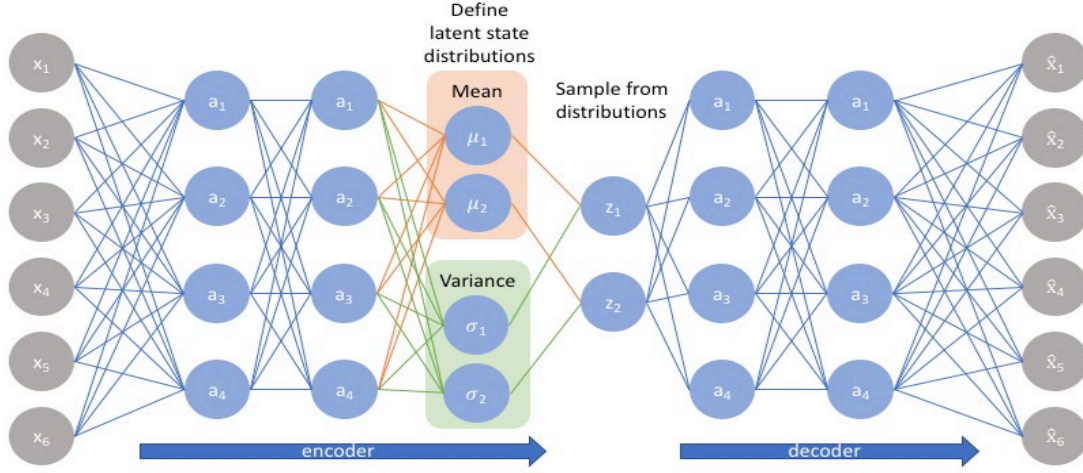
شهدت التطورات الحديثة في التعلم العميق إمكانية تدريب الآلات على التعرف على الأشياء تحت ظروف الرؤية المختلفة. يتم ذلك عن طريق تدريب الآلة على مجموعة كبيرة من الصور، حيث يتم تسمية كل صورة بالكائن المتوافق. تتعلم الآلة التعرف على الأشياء من خلال تحديد الأنماط في البيانات، مما يتيح لها التعميم على الصور الجديدة. [24]

ومع ذلك، فإن تخزين واسترجاع المعلومات البصرية في الآلات ما زالت محدودة مقارنة بالإنسان. تستطيع الآلات التعرف فقط على الأشياء التي تم تدريبها عليها، ولا يمكنها التكيف مع أشياء جديدة أو ظروف رؤية جديدة دون تدريب إضافي.

تحسنت مدى تخزين المعلومات البصرية في الآلات في مجال الاستخدامات الصور المضغوطة. يتم استخدام خوارزميات الضغط لتقليل حجم ملفات الصور، مما يجعلها أسهل في التخزين والنقل. ومع ذلك، يمكن أن يؤدي ضغط الصور إلى فقدان الجودة، مما يجعل من الصعب على الآلات التعرف على الأشياء في الصورة.

للتغلب على هذه المشكلة، طوّر الباحثون خوارزميات ضغط جديدة مصممة خصيصاً لتطبيقات التعلم الآلي. تستخدم هذه الخوارزميات تقنيات مثل التكوين الكمي، وترميز الانتروبيا، وترميز التحويلات لتقليل حجم ملفات الصور وفي نفس الوقت الحد من فقدان الجودة.

ومن طرق الضغط أيضاً، الرموزات التلقائية المتغيرة التي تحاول إيجاد التوزيع الاحتمالية التي تنتمي إليها البيانات، فيتم تدريب هذه النماذج على مجموعة بيانات من الصور وتكون مهمة الآلات في هذه الحالة هي كسر الصور المدخلة وإعادة تجميعها بدلالة المتغيرات الأكثر أهمية. [23]



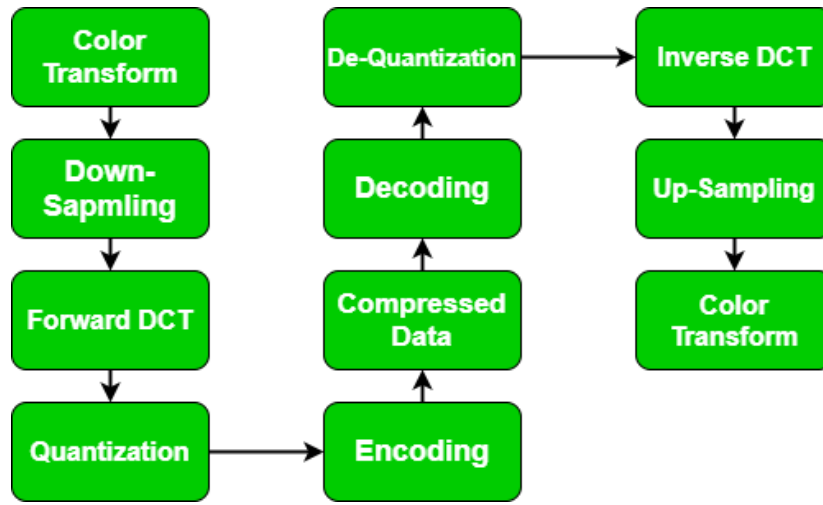
Variational Auto Encoders Architecture19 figure

لخلاصة القول، على الرغم من اختلاف تخزين واسترجاع المعلومات البصرية في الإنسان والآلات بشكل كبير، فإن التقدم الحديث في التعلم العميق وخوارزميات الضغط يجعل من الممكن للآلات التعرف على الأشياء تحت ظروف رؤية مختلفة. ومع ذلك، لا تزال الآلات بحاجة إلى المزيد من التطوير لتتمكن من مواكبة مرونة وقدرة التكيف في النظام البصري البشري.

صيغ وخوارزميات ضغط الصور

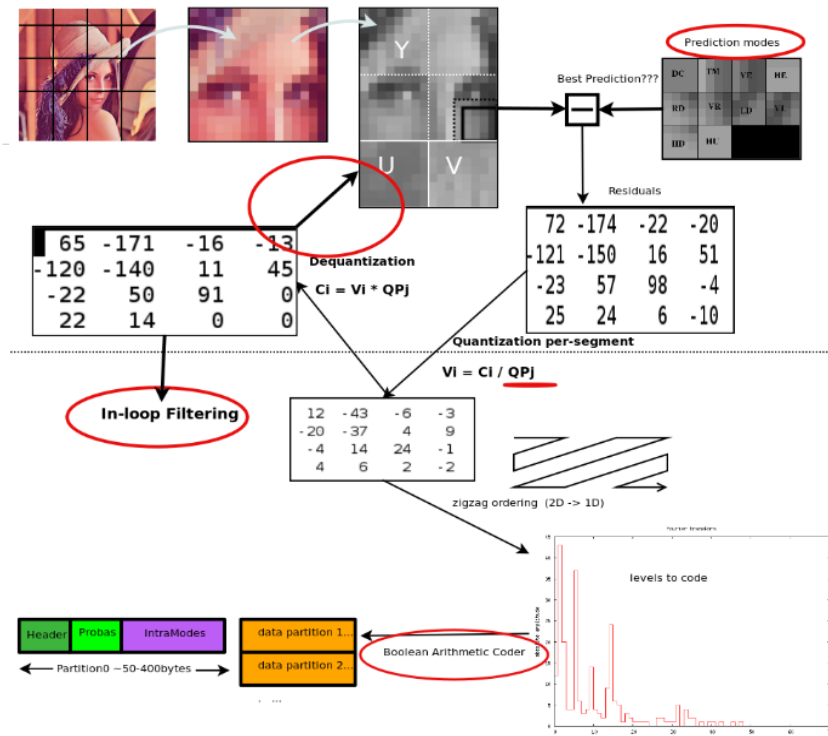
تأتي الصور بصيغ وأحجام مختلفة، ويمكن ضغطها باستخدام خوارزميات مختلفة لتقليل حجمها مع الحفاظ على جودتها البصرية. في العصر الرقمي، الصور شائعة وتستخدم في مجالات مختلفة، بما في ذلك رؤية الحاسوب، وتعلم الآلة، والروبوتات. لذلك، فإن فهم الصيغ وخوارزميات ضغط الصور المختلفة أمر ضروري للمحترفين العاملين في هذه المجالات. [24]

أحد أكثر صيغ الصور شيوعاً هي (JPEG- Joint Photographic Experts Group) التي تم إدخالها في عام 1992. وهي خوارزمية ضغط تفقدية تستخدم التحويل المتسلسل للتجيب (DCT – Discrete Cosine Transform) لتحويل الصورة إلى تمثيل المجال الترددي. يتم تكوين معاملات DCT ، ثم يتم ترميز القيم الناتجة باستخدام نظام الترميز بالطول المتغير. يتم استخدام صيغة JPEG على نطاق واسع في التصوير الرقمي ويتم دعمها بشكل تقريبي من قبل جميع برامج تحرير الصور وعرضها. [24]



JPEG Algorithm20 figure

صيغة الصور الشائعة الأخرى هي صيغة الشبكة المحمولة (PNG) ، والتي تم إدخالها في عام 1996 كبديل لصيغة تبادل الرسومات PNG (GIF). هي خوارزمية ضغط غير تفقدية تستخدم مزيجاً من اثنين من الأساليب المختلفة: خوارزمية Deflate ونظام الترميز Adaptive Huffman. صيغة PNG مناسبة تماماً للصور التي تحتوي على مناطق كبيرة من اللون الموحد، مثل الشعارات والرموز.



PNG Compression Algorithm21 figure

اختيار صيغة الصور وخوارزمية الضغط يعتمد على عوامل مختلفة، مثل الاستخدام المقصود للصورة والمستوى المرغوب من جودة الصورة والتخزين والنطاق الترددي المتاح. يجب على المحترفين في رؤية الحاسوب والتعلم الآلي والروبوتات أن يكونوا على دراية بمختلف صيغ الصور وخوارزميات الضغط لاختيار الخيار الأفضل لحالتهم الاستخدامية المحددة.

خوارزميات الضغط ضرورية لتقليل متطلبات التخزين والنقل للصور الرقمية. تحفظ خوارزميات الضغط غير التلقيدية جميع البيانات الأصلية في الصورة، بينما تتضح البيانات في الخوارزميات التلقيدية لتحقيق حجم ملفات أصغر. يعتمد اختيار خوارزمية الضغط على الاستخدام المقصود للصورة والتخزين والنطاق الترددي المتاح والمستوى المرغوب من جودة الصورة.

3.1 الشبكات العصبونية الالتفافية (Convolutional Neural Networks CNNs)

هي نوع من الشبكات العصبية تم تصميمها خصيصاً لمهام التعرف على الصور والفيديو. تم تقديمها لأول مرة في الثمانينيات من القرن الماضي، لكنها لم تنتشر حقاً حتى أوائل عقد 2010، وذلك بفضل توفر مجموعات بيانات كبيرة، مثل ImageNet، وتطوير وحدات المعالجة الرسومية القوية التي يمكنها تدريب الشبكات الكبيرة بسرعة. [26]

الفكرة الرئيسية وراء CNNs هي تطبيق مرشحات (المعروفة أيضاً باسم الأوزان أو النوى) على الصورة لاستخراج الميزات في مواقع مختلفة. يتم دمج هذه الميزات في طبقات متعاقبة لتشكيل ميزات عالية المستوى، وهكذا، حتى تنتج طبقة الإخراج النهائية مجموعة من التنبؤات لصورة الإدخال.

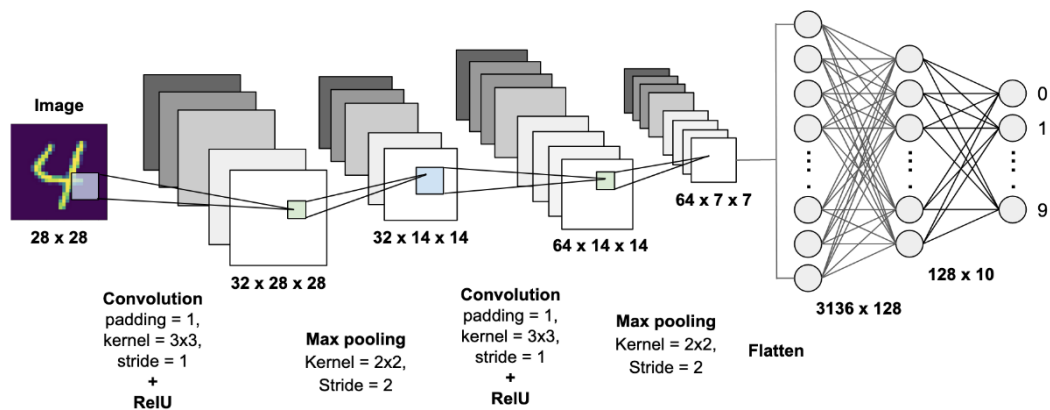


Figure 22 Basic Convolutional Neural Network for Handwritten Digit Classification

تتألف معمارية الشبكة العصبية الالتفافية التقليدية من ثلاثة أنواع من الطبقات: طبقات الالتفاف (Convolutional Layers)، وطبقات الاقتراع (Pooling Layers)، وطبقات متصلة بالكامل (Fully Connected Layers). وتقوم طبقات الالتفاف بتطبيق مرشحات على الصورة المدخلة، في حين تقوم طبقات الاقتراع بتخفيض حجم خرائط السمات التي تنتجها طبقات الالتفاف، وهذا يقلل من التكلفة الحسابية للطبقات التالية. وتستخدم طبقات متصلة بالكامل لإنتاج الناتج النهائي، وعادة ما تأتي متبوعة بتابع تفعيل مثل SoftMax الذي يحول الناتج إلى توزيع احتمالي على الفئات (Probability Distribution). [22]

أحد مزايا الشبكات العصبية الالتفافية (CNN) هو قدرتها على تعلم تمثيلات الميزات تلقائياً من البيانات الخام، دون الحاجة إلى تصميم ميزات باليد (Hand-Engineered Features). وهذا يجعلها قادرة على التكيف مع مجموعة واسعة من مهام التعرف على الصور، بدءاً من كشف الكائنات (Object Detection) والشرائح (Segments) إلى وصف الصور (Image Description) ونقل الأسلوب العصبي (Neural Style) (Transfer).

في السنوات الأخيرة، تم استخدام الشبكات العصبية الالتفافية (CNNs) لتحقيق أداء يصل إلى أحدث الإنجازات على مجموعة متنوعة من معايير التعرف على الصور، بما في ذلك ImageNet و COCO. وقد تم تطبيقها أيضاً على مجالات أخرى، مثل معالجة اللغة الطبيعية والتعرف على الكلام، حيث حققت نتائج مشجعة.

على الرغم من نجاحها، إلا أن الشبكات العصبونية الالتفافية (CNNs) ليست بلا عيوب. فهي تتطلب كميات كبيرة من بيانات التدريب ويمكن أن تكون حساسة للتغيرات الصغيرة في الإدخال مثل الدوران أو التحجيم. وهي عادة ما تكون باهظة التكلفة حسابياً، خاصة عندما تُطبق على الصور أو مقاطع الفيديو ذات الدقة العالية.

في الأقسام التالية، سنستكشف بعض المكونات الرئيسية لشبكات CNN بالتفصيل، بما في ذلك معايير اختيار الشبكة الصحيحة لحل مشكلة البحث المعينة، والمدخلات والطبقات المستخدمة في شبكات CNNs، وتوابع التفعيل Sigmoid و ReLU، وطبقات الاقتراع Pooling Layers، ووصلات التخطي Skip Connections، وطبقات الانحراف المكافئ، وتدريب الشبكات، وتركيب الشبكة المشهورة UNet [13].

معايير اختيار الشبكة العصبية الالتفافية:

اختيار الشبكة العصبية الالتفافية المناسبة لمشكلة بحثية محددة في الاستدلال بالصور أمر حاسم لتحقيق أداء عالي واستخدام فعال للموارد. هناك عدة مقاييس يمكن استخدامها لتقييم مدى ملائمة شبكة CNN لمهمة محددة، بما في ذلك الدقة والسرعة واستخدام الذاكرة.

الدقة Accuracy: هي ربما أهم المقاييس لقياس أداء شبكة CNN. فهي تقيس مدى قدرة الشبكة على التنبؤ بالنتيجة الصحيحة لمدخل محدد. عند تقييم دقة CNN، من المهم استخدام مجموعة بيانات وبروتوكول تقييم مناسبين يعكسان خصائص التطبيق المستهدف. على سبيل المثال، إذا كانت المهمة تتضمن كشف الكائنات في الصور الطبيعية، يجب تقييم الشبكة على مجموعة بيانات تحتوي على مجموعة متنوعة من الكائنات في مشاهد واقعية، بمجموعة متنوعة من الأحجام ونسب الجوانب للكائنات. [26]

السرعة Speed: هي مقياس آخر مهم، خاصة في التطبيقات في الوقت الحقيقي حيث يكون وقت الاستخراج مهماً. تتأثر سرعة شبكة CNN بعدة عوامل، بما في ذلك حجم الشبكة وعدد المعلمات وتعقيد العمليات والأجهزة المستخدمة للاستخراج. لتقييم سرعة CNN ، يقيس الباحثون عادة الوقت الذي يستغرقه المعالج لمعالجة صورة مدخل واحدة على منصة عتاد محددة.

استخدام الذاكرة Memory Usage: هو مقياس ثالث يمكن استخدامه لتقييم مدى ملائمة CNN لمشكلة البحث المحددة. يتم تحديد استخدام الذاكرة بحجم الشبكة وكمية الذاكرة المطلوبة لتخزين التفعيلات الوسيطة أثناء الاستخراج. تميل الشبكات الكبيرة التي تحتوي على العديد من الطبقات والمعاملات إلى تتطلب مزيداً من الذاكرة من الشبكات الأصغر، ويمكن أن تصبح هذه المشكلة عاملاً محدداً على الأجهزة المحدودة من حيث الموارد.

المقاييس الأخرى التي يمكن استخدامها لتقييم أداء شبكات CNN تتضمن كفاءة الطاقة والمتانة أمام الهجمات العدائية والتفسيرية. تعد كفاءة الطاقة مهمة للأجهزة التي تعمل بالبطارية، بينما تعد المتانة أمام الهجمات العدائية مهمة لتطبيقات الأمان. بينما تعد التفسيرية مهمة لفهم كيفية اتخاذ الشبكة لتوقعاتها وتحديد الانحيازات أو الأخطاء المحتملة.

في الواقع، غالباً ما يحتاج الباحثون إلى موازنة هذه المقاييس المختلفة اعتماداً على المتطلبات الخاصة لمشكلة البحث المحددة. على سبيل المثال، إذا كانت المهمة هي كشف الأمراض النادرة من الصور الطبية، فقد تكون الدقة هي المقياس الأكثر أهمية، بينما قد يكون استخدام الذاكرة أقل قلقاً. ومن ناحية أخرى، إذا كانت المهمة هي تنفيذ تحليلات فيديو في الوقت الحقيقي على جهاز محمول، فقد يكون السرعة واستخدام الذاكرة هما أكثر أهمية من الدقة.

بالنسبة لتطبيقنا وبسبب محدودية الموارد الحسابية، فقد قمنا باختيار الشبكة على أساس الاستهلاك الأمثل للموارد الحسابية المتاحة.

في المجمل، يتضمن اختيار الشبكة العصبية الالتفافية المناسبة لمشكلة بحث محددة تقيماً لعدة مقاييس، بما في ذلك الدقة والسرعة واستخدام الذاكرة. يحتاج الباحثون إلى توازن هذه المقاييس بناءً على متطلبات التطبيق المحدد والموارد المتاحة.

دخول الشبكات العصبية الالتفافية:

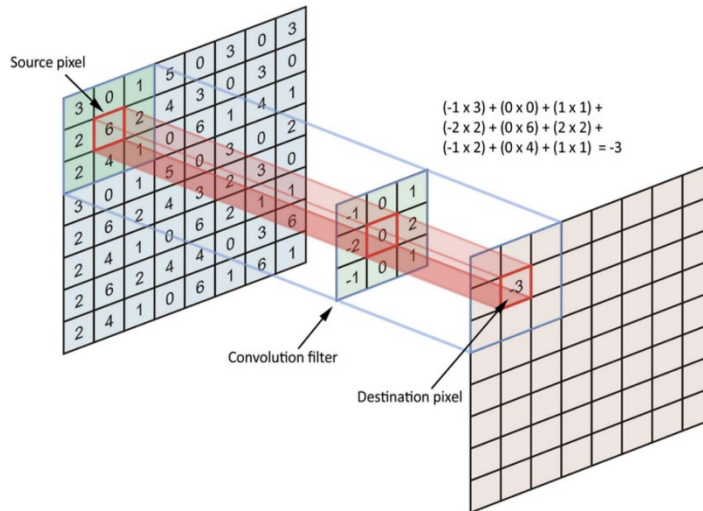
تتكون الشبكات العصبية الالتفافية عادةً من عدة طبقات، يقوم كل منها بعملية معينة على البيانات المدخلة. المدخلات لـ CNN عادةً ما تكون صورة ثنائية الأبعاد، على الرغم من أنها يمكن أن تكون حجماً ثلاثي الأبعاد (على سبيل المثال، الفيديو). هدف CNN هو تعلم تسلسل ميزات تمثل البيانات المدخلة بطريقة مفيدة للمهمة المستهدفة.

الصورة المدخلة:

الطبقة الأولى في CNN هي الطبقة المدخلة، والتي تستقبل الصورة المدخلة الخام. يتم تمثيل الصورة المدخلة عادةً على شكل مصفوفة ثلاثية الأبعاد بأبعاد بالشكل التالي (الارتفاع، العرض، القنوات)، حيث تمثل الارتفاع والعرض أبعاد الصورة، والقنوات تمثل قنوات الألوان (مثل قنوات الألوان الأساسية الثلاث -أحمر، أخضر، أزرق- RGB أو صور التقييس الرمادي Grayscale Images) على سبيل المثال، ستمثل صورة RGB بأبعاد 224×224 على شكل مصفوفة من الحجم (224, 224, 3).

قبل إدخال الصورة إلى CNN، غالباً ما يتم معالجتها مسبقاً لتطبيع قيم البكسل وتقليل تأثيرات التباين في الإضاءة. تتضمن الخطوات الشائعة للمعالجة المسبقة طرح قيمة البكسل المتوسطة والقسم على الانحراف المعياري، أو ضمن أغلب التطبيقات العملية، قسمة قيم جميع البكسلات على أقصى قيمة ممكنة للبكسل (255 في حال استخدام نمط البيانات Unsigned Int 8)

الطبقة الالتفافية Convolutional Layer: الطبقة الأكثر أهمية في CNN هي الطبقة الترابطية، والتي تقوم بعملية الترابط على البيانات المدخلة. تشمل عملية الترابط سحب فلتر صغير (يسمى أيضاً النواة أو الوزن) عبر الصورة المدخلة وحساب حاصل الضرب الالتفافي بين الفلتر والشق المدخل المقابل في كل موقع.



Convolution Operation23 figure

مخرجات الطبقة الالتفافية هي مجموعة من خرائط الميزات، والتي تمثل وجود أو عدم وجود معالم معينة في مواقع مختلفة في الصورة المدخلة. يتم تحديد عدد خرائط الميزات بواسطة عدد الفلاتر في الطبقة الترابطية، ويتم إنشاء كل خريطة ميزات بواسطة فلتر مختلف. يتم تعلم كل فلتر بواسطة CNN أثناء التدريب للكشف عن نوع معين من الميزات، مثل الحواف، والزوايا، والبقع.

حجم الفلاتر والخطوة (أي، المسافة بين مواقع الفلاتر المجاورة) هي معاملات فورية يمكن ضبطها للتحكم في حجم خرائط الميزات الناتجة وحقل الاستجابة للشبكة. حجم الفلتر الأكبر والخطوة الأكبر يؤديان إلى حجم إخراج أصغر وحقل استقبال أكبر، بينما يؤدي حجم الفلتر الأصغر والخطوة الأصغر إلى حجم إخراج أكبر وحقل استقبال أصغر.

لتلخيص، المدخل إلى CNN عادةً ما تكون الصورة ثنائية البعد ممثلة بمصفوفة ثلاثية الأبعاد، والطبقة الأكثر أهمية في CNN هي الطبقة الالتفافية، التي تنفذ عملية الترابط على البيانات المدخلة وتولد مجموعة من خرائط الميزات. حجم الفلاتر والخطوة هي معاملات فورية يمكن ضبطها للتحكم في حجم خرائط الميزات الناتجة وحقل الاستجابة للشبكة.

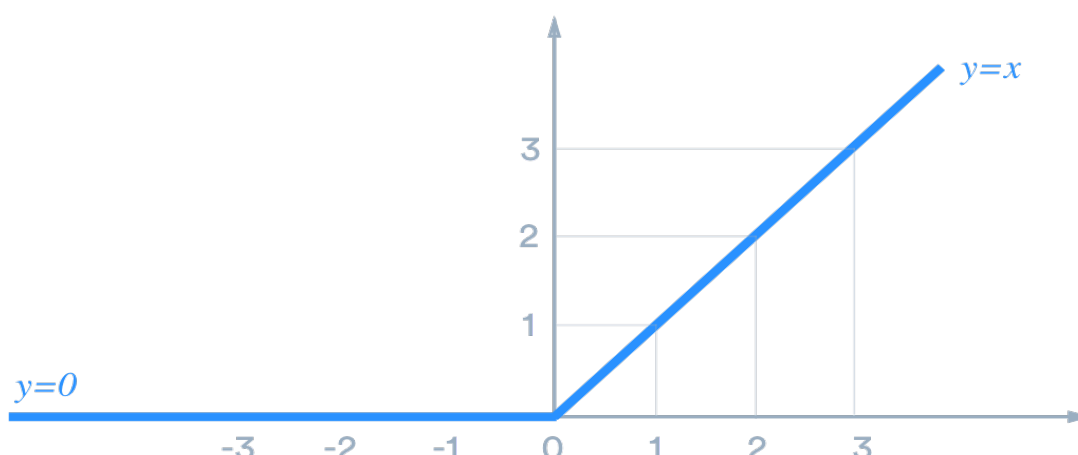
وحدة الانحدار الخطي المصححة (ReLU):

هي واحدة من أشهر توابع التصنيف المستخدمة في الشبكات العصبية الالتفافية CNNs تعتبر وظيفة ReLU بسيطة وفعالة من حيث الحسابات، مما يجعلها مثالية للشبكات العصبية العميقة التي تتطلب عدداً كبيراً من المعلمات. [1]

يتم تعريف وظيفة ReLU على النحو التالي:

$$F(x) = \max(0, x)$$

حيث x هو المدخل إلى الدالة، و $f(x)$ هو الإخراج. تعيد وظيفة ReLU قيمة المدخل إذا كانت إيجابية، والصفر في غير ذلك. هذا يعني أن ReLU هي وظيفة عتبية تقوم بتعيين القيم السلبية على الصفر.



تابع التفعيل (-ReLU Rectified Linear Unit) واحداً من التوابع الأكثر استخداماً في الشبكات العصبية الالتفافية (Convolutional Neural Networks -CNNs). ويعتبر تابع التفعيل ReLU بسيط وفعال حسابياً، مما يجعله مثالياً للشبكات العصبية العميقة التي تتطلب عدداً كبيراً من المعاملات.

تابع التفعيل ReLU لها عدة مزايا على الوظائف الأخرى مثل Sigmoid و Tanh واحدة من أهم المزايا هي أنها فعالة حسابياً، حيث تتطلب فقط عملية مقارنة بسيطة وعملية max، هذا يجعلها أسرع في الحساب من Sigmoid و Tanh التي تتطلب عمليات رياضية أكثر تعقيداً.

ومن المزايا الأخرى لـ ReLU هو أنها يمكن أن تساعد في التخفيف من مشكلة اختفاء المشتق (Vanishing Gradient Problem) التي يمكن أن تحدث في الشبكات العميقة. تحدث مشكلة اختفاء المشتق عندما تصبح قيم المشتقات صغيرة جداً بمجرد أن تمر من خلال الشبكة، مما يمكن أن يؤدي إلى بطء التعلم أو منعه. ويمكن لـ ReLU أن يساعد في منع القيم التدريجية من الانخفاض بشكل كبير لأنها لا تنتشع، وبالتالي لا تقترب من الحد الأقصى أو الحد الأدنى.

وعلاوة على ذلك، يمكن أن يساعد ReLU على تحسين قلة كثافة الشبكة، حيث يقوم بتحويل عدد كبير من التنشيطات إلى الصفر. وهذا يمكن أن يؤدي إلى استخدام أكثر كفاءة لمعاملات الشبكة، إذ يحد التكرار في الشبكة.

على الرغم من مزاياه، لدى ReLU أيضاً بعض القيود. واحدة من أهم القيود هي أنه يمكن أن يعاني من مشكلة (Dying ReLU)، حيث تصبح بعض العصبونات غير فعالة بشكل دائم ولا تساهم في الخرج. ويمكن أن يحدث ذلك عندما يكون دخل العصبون سلبياً بشكل متكرر، حيث يصبح معدل التغيير التدريجي الخاص بالعصب صفرًا ويتوقف التعلم. وللتخفيف من هذه المشكلة، تم اقتراح العديد من الإصدارات من ReLU، مثل (Leaky ReLU) و (Parametric ReLU) والتي تضيف منحنى قليلاً أو معلمة يمكن تعلمها إلى الجانب السلبي من الدالة.

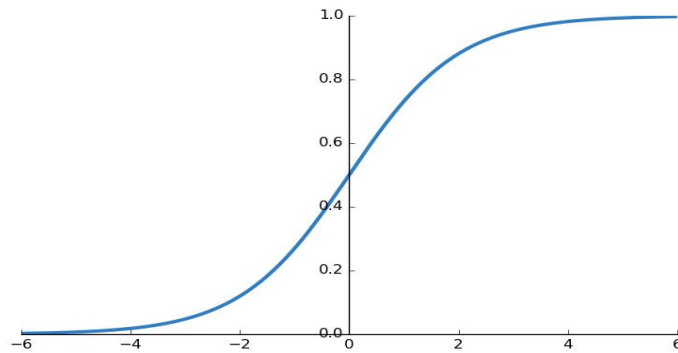
باختصار، وحدة الانحراف الخطي المصححة (ReLU) هي تابع تفعيل شائع الاستخدام في الشبكات العصبية الالتفافية (CNNs) تعتبر وحدة ReLU فعالة حسابياً وتساعد على التخفيف من مشكلة اختفاء المشتق (Vanishing Gradient Problem) الذي يمكن أن يحدث في الشبكات العصبية العميقة، ويمكن أن تحسن من قلة كثافة الشبكة. ومع ذلك، يمكن أن تعاني من مشكلة (Dying ReLU) مما أدى إلى تطوير متغيرات مثل (Leaky ReLU) و (Parametric ReLU).

تابع التفعيل Sigmoid:

تابع التفعيل سيغمويد (Sigmoid) هو تابع غير خطي Non-Linear Function مستخدم على نطاق واسع في تعلم الآلة والتعلم العميق. وغالباً ما يستخدم في الشبكات العصبية الالتفافية (Convolutional Neural Networks) بسبب قدرته على إنتاج إخراج احتمالي [1]. يعرف تابع Sigmoid بالعلاقة التالية:

$$F(x) = \frac{1}{(1 + e^{-x})}$$

حيث x هو المدخل إلى الدالة و $f(x)$ هو الإخراج. ترتبط وظيفة السيغمويد بأي قيمة مدخلة بقيمة بين 0 و 1. وهذا يجعلها مثالية لمهام التصنيف حيث نريد تنبؤ احتمالية فئة معينة.



Sigmoid Activation Function25 figure

يتمتع تابع Sigmoid بالعديد من المزايا، فهو تابع مستمر يمكن اشتقاقه على كامل مجاله مما يجعله مناسباً لتقنيات التحسين (Optimization Techniques) مثل التدرج النزولي (Gradient Descent). بالإضافة إلى ذلك، فإن مستقر التابع الواقع في المجال $[0, 1]$ يساعد في إنتاج توزيع احتمالات على الفئات.

ومع ذلك، فإن هذا التابع يمتلك بعض العيوب. واحدة من العيوب الرئيسية هي أنها يمكن أن تعاني من مشكلة اختفاء المشتقات (Vanishing Gradients). تحدث هذه المشكلة عندما يصبح مشتق التابع Sigmoid صغير جداً عندما تتحرك القيم المدخلة بعيداً عن الصفر. هذا يمكن أن يجعل تدريب الشبكات العصبية العميقة باستخدام وظائف تابع التفعيل Sigmoid أمراً صعباً، حيث يمكن أن يصبح التدرج صغيراً جداً للسماح بتحديثات ذات مغزى للأوزان، لذلك غالباً ما نرى هذا التابع في الطبقة أو الطبقات الأخيرة (طبقات الخرج) حيث أنه لا مفر من استخدامه (أو مثيل له) من أجل عمليات التصنيف.

عيب آخر لوظيفة السيغمويد هو عدم كونها مركزة على الصفر. هذا يمكن أن يؤدي إلى مشاكل أثناء التدريب حيث يمكن أن يتسبب في تحيز التدرجات نحو اتجاهات معينة، مما يؤدي إلى تباطؤ التقارب أو الحلول غير الأمثل.

على الرغم من هذه القيود، تظل وظيفة التنشيط السيغمويد خياراً شائعاً لبعض المهام، مثل مشكلات التصنيف الثنائي حيث نريد التنبؤ بقيمة احتمالية بين 0 و 1.

وبشكل ملخص، تعتبر وظيفة التنشيط السيغمويد دالة غير خطية تستخدم على نطاق واسع في تعلم الآلة والتعلم العميق، وخاصة في الشبكات العصبية الالتفافية (CNN) لمهام التصنيف. تنتج وظيفة السيغمويد إخراجاً احتمالياً محصوراً بين 0 و 1. على الرغم من وجود بعض المزايا مثل السلاسة والإخراج الاحتمالي، فإنها تمتلك بعض القيود مثل مشكلة اختفاء التدرج وعدم التركيز على الصفر.

طبقة الاقتراع (Pooling Layer):

هي مكون مهم في الشبكات العصبية الالتفافية (CNNs) المستخدمة لاستخراج الميزات. إنها تقنية تستخدم لتقليل الأبعاد المكانية لحجم المدخلات، مع الحفاظ على السمات الأكثر أهمية.

تعمل طبقات الاقتراع على خرج طبقة الالتفاف، والتي تعد خريطة الميزات. ويتمثل هدف طبقة الاقتراع في التخفيض في حجم خريطة الميزات عن طريق تحجيمها، مع الحفاظ على السمات الأكثر أهمية. وهذا يساعد في تقليل الحسابات اللازمة لمعالجة خريطة الميزات في الطبقات التالية، مع تقليل خطر الاستحواذ الزائد (Overfitting). [26]

هناك أنواع عديدة من طبقات الاقتراع، بما في ذلك الاقتراع الأقصى (Max Pooling) وفيه يتم اختيار الميزة ذات قيمة التفعيل الأعلى في كل نافذة للاقتراع (Sliding Window)، والاقتراع المتوسط (Average Pooling) وفيه يتم حساب المتوسط الحسابي للميزات الموجودة في نافذة الاقتراع الحالية، وأيضاً لدينا اقتراع التطبيع L2 (L2-Norm Pooling) وفيه يتم حساب قيمة التطبيع L2 (L2-Norm) للميزات الموجودة في نافذة الاقتراع، يعد الاقتراع الأقصى هو النوع الأكثر استخداماً من الطبقات المجمعة. يعمل عن طريق تقسيم خريطة الميزات المدخلة إلى مناطق مستطيلة غير متداخلة وإخراج القيمة القصوى في كل منطقة. يتم تحديد حجم المناطق بواسطة نافذة الاقتراع، التي تكون عادةً منطقة مربعة أو مستطيلة صغيرة. يكون إخراج طبقة الاقتراع الأقصى هو خريطة الميزات المخفضة التي تكون أصغر من خريطة الميزات المدخلة.

واحدة من المساوئ المحتملة لطبقات الاقتراع هي أنها يمكن أن تؤدي إلى فقدان بعض المعلومات المكانية. نظراً لأن التجميع ينطوي على تخفيض أبعاد خريطة الميزات، يمكن فقدان بعض التفاصيل المكانية (Spatial Details)

في العملية. يمكن التخفيف من ذلك عن طريق استخدام خطوات أصغر في الطبقات الالتفافية أو باستخدام تقنيات أخرى مثل الالتفافات المتباعدة.

لخلاصة الفقرة، طبقة الاقتراع هي عنصر حاسم في أداء الشبكات العصبونية الالتفافية المستخدمة لاستخراج الميزات. إنها تساعد على تقليل الأبعاد المكانية لخريطة الميزات الإدخالية، مع الاحتفاظ بالميزات الأكثر أهمية. هناك عدة أنواع من طبقات الاقتراع، بما في ذلك الاقتراع الأقصى والاقتراع المتوسط و اقتراع L2-Norm، وعلى الرغم من أهميتها وفائدتها، يجب اتخاذ الحيلة والحذر للحد من فقدان المعلومات المكانية التي يمكن أن تحدث خلال عملية التجميع.

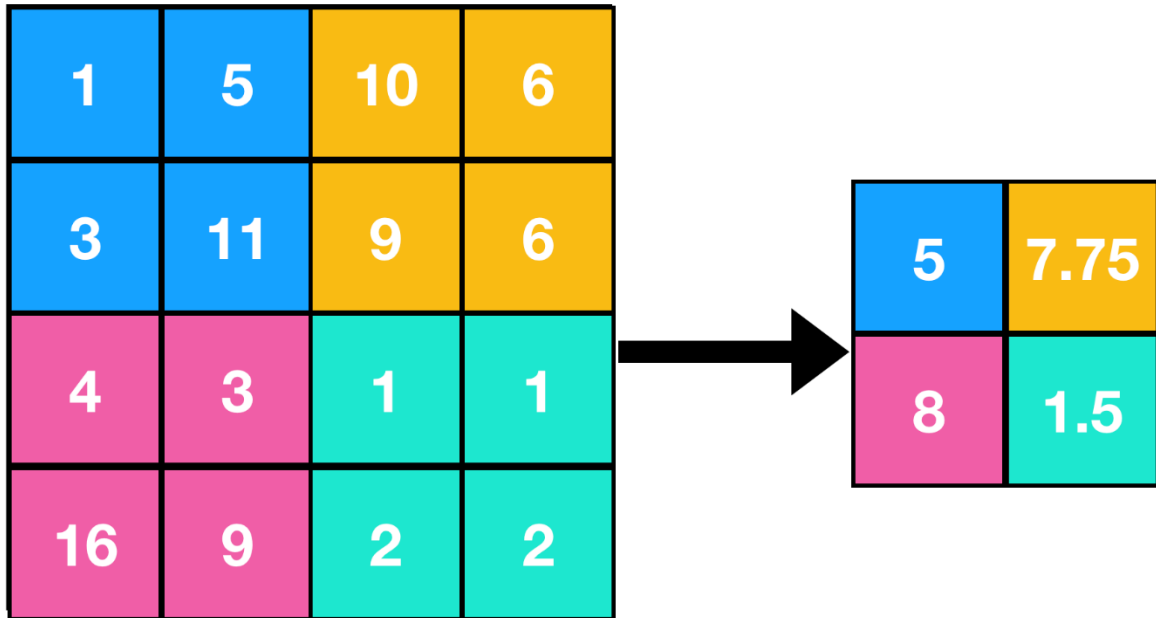
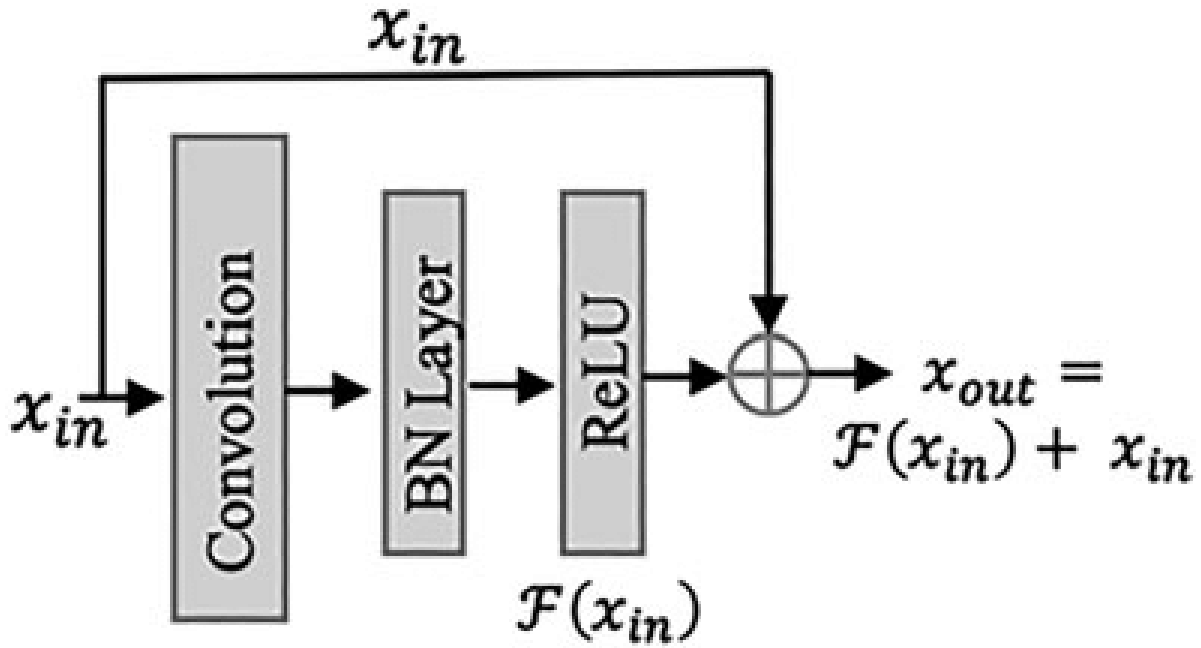


figure 26 Average Pooling

وصلات التخطي *Skip Connections*:

تعتبر وصلات التخطي *Skip Connections*، المعروفة أيضاً باسم وصلات البواقي *Residual Connections*، نوعاً من التقنيات الكلاسيكية التي يتم استخدامها في الشبكات العصبية العميقة. وقد تم تقديمها لأول مرة في بنية *ResNet*، التي حققت أداءً متفوقاً في العديد من المهام الرئيسية لرؤية الحاسوب، بما في ذلك تصنيف الصور والكشف عن الكائنات، حيث استطاع فريق الباحثين الذين اخترعوا التقنية من الفوز بعدة مسابقات بمعايير مختلفة. [26]

في الشبكة العصبية، تسمح وصلات التخطي للمعلومات بالانتقال من طبقة إلى أخرى دون الذهاب من خلال الطبقات الوسيطة. ويمكن أن يكون هذا مفيداً لعدة أسباب. أولاً، يمكن أن يساعد في التخفيف من مشكلة اختفاء المشتقات Vanishing Gradients، والتي يمكن أن تحدث في الشبكات العميقة عندما تصبح المشتقات صغيرة جداً لتحديث الأوزان بشكل فعال. من خلال السماح للمعلومات بالتجاوز من الطبقات الوسيطة، يمكن لوصلات التخطي المساعدة في الحفاظ على قيم غير صفيرية للمشتقات وتحسين الأداء العام للشبكة، وتعد هذه التقنية السبب الأساسي الذي ساهم في بدء الأبحاث لشبكات أكثر عمقاً، حيث أن الكثير من الشبكات العصبية العميقة التي يتم إجراء الأبحاث عنها اليوم -والتي يكون عمقها من مرتبة مئات الطبقات- تستخدم وصلات التخطي أو تقنيات مشتقة منها لدعم قيم المشتقات في الطبقات العميقة.



Basic Identity Skip Connection27 figure

بالإضافة إلى ذلك، يمكن أن تساعد وصلات التخطي على تحسين دقة الشبكة عن طريق توفير اتصالات أكثر مباشرة بين المدخلات والمخرجات. ويمكن أن يكون هذا مفيداً بشكل خاص للمهام التي تتمتع المدخلات والمخرجات فيها بترابط قوي، مثل تجزئة الصور أو ترجمة الصورة إلى صورة أخرى.

هناك العديد من أنواع طبقات التخطي، بما في ذلك وصلات تخطي الهوية Identity Skip Connection، وصلات تخطي الإسقاط Projection Skip Connection، ووصلات التخطي الكثيفة Dense Skip Connections. وتعتبر وصلات تخطي الهوية هي النوع الأبسط، وتنطوي على إضافة المدخلات إلى مخرج طبقة الالتفاف بعد إعادة تقييسها لتمكين دمج المصفوفات، ويمكن أن تساعد هذه الطريقة في الحفاظ على الدقة المكانية لخريطة المعالم وتحسين أداء الشبكة.

أما بالنسبة لوصلات تخطي الإسقاط Projection Skip Connection، فتتضمن استخدام طبقة تحويل تكاملي للترددات لإسقاط المدخلات على نفس الأبعاد كمخرج الشبكة. ويمكن أن تكون هذه الطريقة مفيدة في الحالات التي تختلف فيها الدقة المكانية أو أبعاد المعالم بين المدخلات والمخرجات.

وتعد وصلات الهروب الكثيفة هي نوع متقدم أكثر من اتصالات المرور المتخطية، وتنطوي على ربط كل طبقة بكل طبقة أخرى في الشبكة. ويمكن أن تساعد هذه الطريقة في تحسين تدفق المعلومات بين الطبقات وتقليل مخاطر حدوث الانسداد في معلومات الشبكة.

أصبحت وصلات التخطي عنصراً مهماً في العديد من بنى الشبكات العصبية العميقة الحديثة، بما في ذلك ResNet و DenseNet و U-Net. وقد ثبت أنها فعالة لمجموعة واسعة من مهام رؤية الحاسوب، بما في ذلك التصنيف الصوري، وكشف الكائنات، والتجزئة الدلالية، والترجمة من صورة إلى صورة.

وبإيجاز، فإن وصلات التخطي هي نوع من وصلات الاختصار Shortcut Connections التي تستخدم في شبكات العصبية العميقة، ويمكن أن تساعد في تخفيف مشكلة اختفاء المشتقات Vanishing Gradients، وتحسين دقة الشبكة، وتوفير اتصالات مباشرة أكثر بين المدخلات والمخرجات. وتتضمن أنواعاً مختلفة من وصلات الهروب، بما في ذلك وصلات هروب الهوية، وصلات هروب الإسقاط، وصلات التخطي الكثيفة، وأصبحت عنصراً مهماً في العديد من بنى الشبكات العصبية العميقة الحديثة.

تدريب الشبكات العصبونية الالتفافية Training Convolutional Neural Networks:

عملية تدريب الشبكات العصبونية الالتفافية Convolutional Neural Networks تتضمن عادة الخطوات التالية:



CNN Training Setup28 figure

1. إعداد البيانات **Data Preparation**: الخطوة الأولى في تدريب CNN هي إعداد البيانات. وقد يتطلب هذا جمع البيانات وتنظيفها، وتغيير حجم الصور، وتقسيم البيانات إلى مجموعات التدريب والتحقق والاختبار (Training, Testing, Validation).

2. اختيار النموذج **Model Selection**: بمجرد أن تم إعداد البيانات، فإن الخطوة التالية هي اختيار بنية CNN مناسبة للمهمة المطلوبة. وقد يتضمن ذلك اختيار نموذج مدرب مسبقاً مثل VGG أو ResNet أو Inception، أو تصميم نموذج مخصص من البداية.

3. تابع الكلفة **Loss Function**: يعد تابع الكلفة مكوناً رئيسياً في عملية التدريب، حيث يقوم بقياس الفرق بين التصنيفات المتوقعة والفعلية لكل صورة في مجموعة التدريب. وتشمل مجموعة توابع الكلفة الأكثر استخداماً في تدريب الشبكات العصبونية الالتفافية خسارة الإنتروبيا القاطعة للتصنيف متعدد الفئات Categorical Cross-Entropy Loss for Multi-Class Classification، وخسارة الإنتروبيا الثنائية للتصنيف الثنائي Binary Cross-Entropy for Binary Classification، ومتوسط مربعات الخط Mean Squared Error لمهام الانحدار Regression Tasks.

4. خوارزمية التحسين **Optimization Algorithms**: تكون خوارزمية التحسين مسؤولة عن تحديث معلمات النموذج لتقليل قيمة تابع الكلفة المستخدم خلال التدريب. وتشمل الخوارزميات الشائعة المستخدمة في CNNs نزول التدرج العشوائي (Stochastic Gradient Descent (SGD)) و تقدير اللحظة التكيفية (Adaptive Moment Estimation (Adam)).

5. ضبط الهايبرباراميتر **Hyperparameter Tuning**: يمكن تحسين أداء النموذج في النهاية من خلال ضبط الهايبربارامترات، مثل معامل سرعة التعلم Learning Rate وحجم الدفعة Batch Size وعدد الدورات التدريبية Epochs وقوة التنظيم Regularization Strength. يمكن القيام بذلك باستخدام البحث الشبكي Grid Search أو البحث العشوائي Random Search، أو باستخدام خوارزميات أكثر تقدماً مثل التحسين البايزي Bayesian Optimization أو الخوارزميات الجينية Genetic Algorithms.

لمراقبة تقدم عملية التدريب، يمكن استخدام العديد من المقاييس، بما في ذلك تابع الكلفة Loss Function والدقة Accuracy والتذكر Recall، يمكن حساب هذه المقاييس على مجموعات التدريب والتحقق والاختبار لتقييم أداء النموذج على بيانات مختلفة.

بالإضافة إلى المكونات الأساسية لتدريب CNN، تم اقتراح العديد من التقنيات لتحسين أداء النموذج، بما في ذلك:

1. **تكبير البيانات Data Augmentation:** يتضمن تكبير البيانات إنشاء بيانات تدريب جديدة عن طريق

تطبيق تحويلات عشوائية على الصور الأصلية، مثل الدوران Rotation والانزياح Translation والتطبيع Normalization. يمكن أن يساعد هذا في منع الإفراط في التعلم Overfitting وتحسين تعميم النموذج Generalization.

2. **نقل التعلم Transfer Learning:** يتضمن نقل التعلم استخدام نموذج CNN مدرب مسبقاً على مجموعة

بيانات كبيرة، مثل ImageNet، كنقطة انطلاق لمهمة جديدة. يمكن ضبط النموذج المدرب مسبقاً على المهمة الجديدة أو استخدامه كمستخرج للميزات الثابت Feature Extractor.

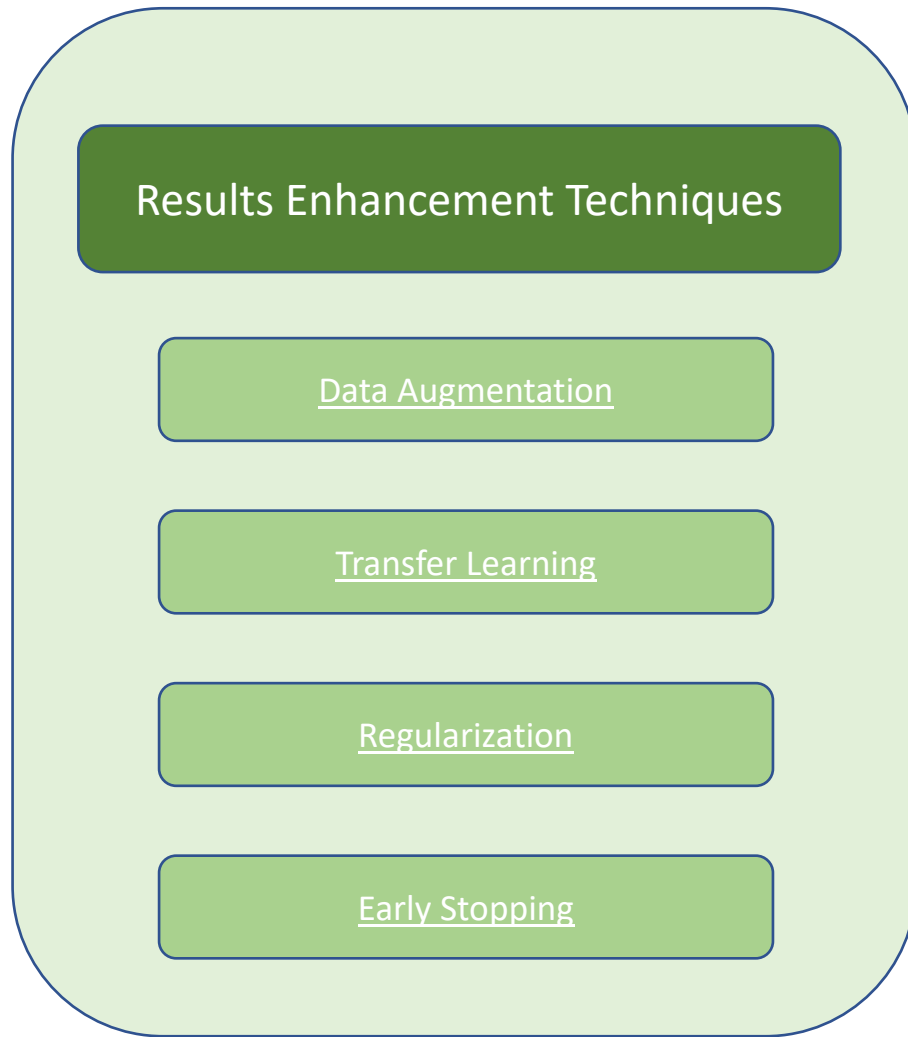
3. **التنظيم Regularization:** يمكن استخدام تقنيات التنظيم، مثل التنظيم L1 (L1-Norm) أو التنظيم L2

(L2-Norm) أو تطبيع الدفعات (Batch Normalization)، لمنع إفراط التعلم Overfitting وتحسين تعميم النموذج Generalization.

4. **التوقف المبكر Early Stopping:** يتضمن التوقف المبكر مراقبة قيمة تابع الكلفة Loss Function

بالنسبة لبيانات التحقق خلال عملية التدريب وإيقاف عملية التدريب عندما تبدأ الخسارة في الارتفاع، مما يشير إلى أن النموذج يعاني من التعلم الزائد Overfitting.

في الخلاصة، يتضمن تدريب شبكات التعلم العميق للتعرف على الصور (CNN) عدة مكونات رئيسية، بما في ذلك إعداد البيانات، اختيار النموذج، اختيار تابع الكلفة المناسب، خوارزمية التحسين، وضبط الهايبربارامترات. باستخدام هذه المكونات بالاشتراك مع التقنيات المتقدمة مثل تعزيز البيانات، ونقل التعلم، والتعويض، يمكن تحقيق أداء رائع في مجموعة واسعة من مهام رؤية الحاسوب.



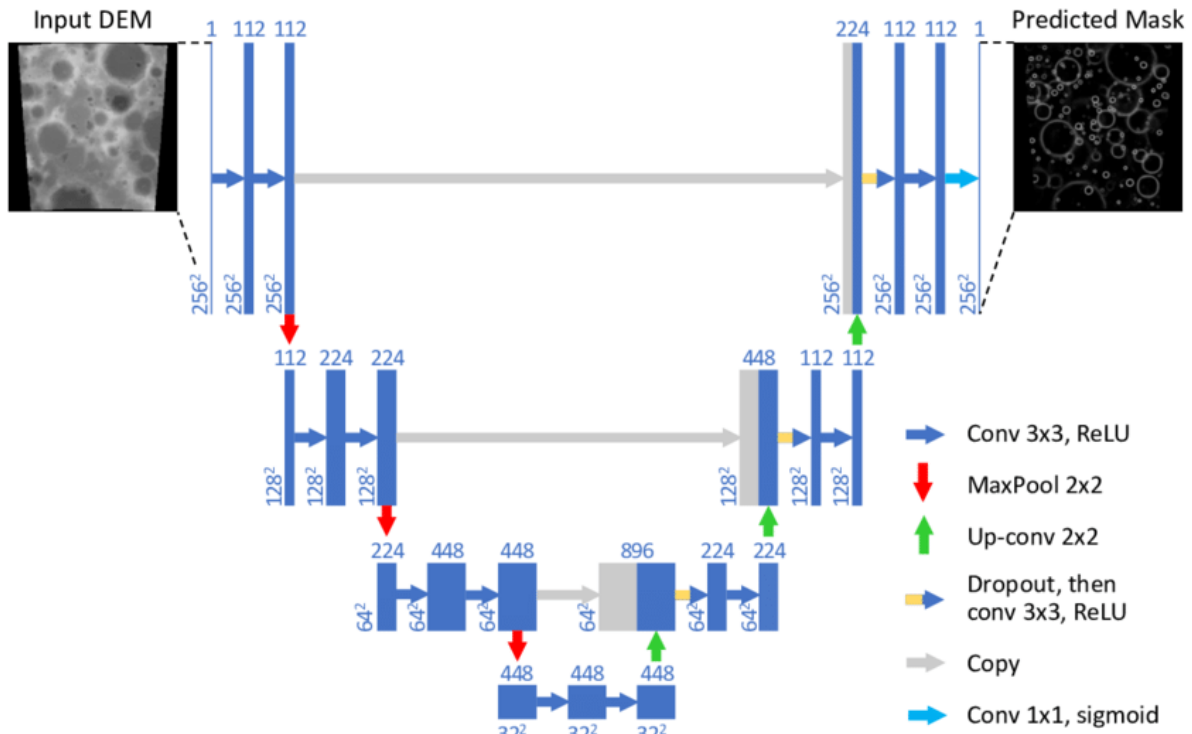
Results Enhancement Techniques29 figure

بنية الشبكة الشهيرة *U-Net*:

في السنوات الأخيرة، تم استخدام نماذج التعلم العميق على نطاق واسع في مجال تجزئة الصور الطبية Medical Images Segmentation. ومن بين هذه النماذج، أصبح U-Net واحداً من أشهر النماذج، وتم استخدامه على نطاق واسع لتطبيقات التصوير الطبي المختلفة.

U-Net هو بنية للشبكات العصبية الالتفافية تم تقديمها من قبل Ronneberger وآخرين في عام 2015 لتجزئة الصور الحيوية. تم تصميمه خصيصاً لمهام تحليل الصور الطبية، مثل تجزئة الخلايا والأنسجة والأعضاء في الصور من المجهرات وغيرها من أنواع صور التصوير الطبي. يأتي اسم "U-Net" من شكل البنية الذي هو متماثل وعلى شكل حرف U. [13]

تتكون بنية U-Net من جزئين رئيسيين: المسار الترميزي (Encoder) والمسار الفكيكي (Decoder). يستخدم المسار الترميزي لالتقاط سياق الصورة المدخلة عن طريق تطبيق عدة عمليات تحويل وتجميع، والتي تقلل من الدقة المكانية Spatial Accuracy لخرائط المعالم Feature Maps مع زيادة عددها. يستخدم المسار التفكيكي، بالمقابل، لاستعادة الدقة المكانية الأصلية للصورة المدخلة عن طريق تطبيق عدة عمليات رفع الدقة والتحويل.



U-Net Architecture30 figure

يتكون المسار الترميزي من عدة طبقات، حيث ينفذ كل منها عمليتي التفاف بمقاس 3×3 متتاليتين، تليها وظيفة تفعيل خطية (ReLU) ثم عملية اقتراح بالقيمة القصوى (Max Pooling) بمقاس 2×2 يتم زيادة عدد خرائط المعالم في كل طبقة تدريجياً في حين يتم تقليل الدقة المكانية. يؤدي هذا العملية إلى تمثيل معالم عالية المستوى للصورة المدخلة.

يتكون المسار التمددي من عدة طبقات، حيث ينفذ كل منها عملية تحويل رفع الدقة متبوعة بالاتحاد مع خريطة المعالم المقابلة من المسار الانقباضي. ثم يتم تطبيق عمليتي تحويل بمقاس 3×3 متتاليتين، تليها وظيفة تفعيل خطية (ReLU). يسمح هذا العملية بإعادة استعادة الدقة المكانية للصورة المدخلة مع الحفاظ على المعلومات الدلالية المستخرجة في المسار الترميزي.

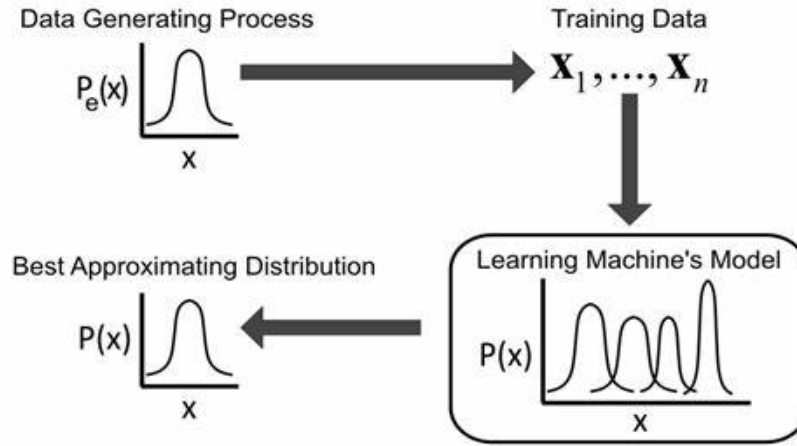
بالإضافة إلى المسار الترميزي والمسار التفكيكي، يتضمن U-Net أيضاً وصلات تخطي بين الطبقات المقابلة في المسار الترميزي والمسار التفكيكي. تسمح هذه الاتصالات باستخدام المعالم عالية المستوى المستخرجة في المسار الترميزي لتحسين التفصيل في المسار التفكيكي.

أصبح نموذج U-Net شائعاً في مجتمع التصوير الطبي بسبب أدائه المتميز في مهام التجزئة المختلفة. وقد أظهرت النتائج الأولية أن النموذج يحقق نتائج عالية الجودة على عدة مجموعات بيانات لصور طبية، بما في ذلك تحدي تتبع الخلايا ISBI وتحدي تجزئة البروستات MICCAI 2015.

وبشكل عام، فإن نموذج U-Net هو تصميم شبكة عصبونية التلافيفية قوية أصبحت خياراً رائجاً للعديد من مهام تجزئة الصور الطبية. وتمكن هذا التصميم من التقاط سياق الصورة المدخلة مع الحفاظ على الدقة المكانية لتجزئة الإخراج بفضل التصميم الرمزي المتناظر والمستهلك من الحرف (U). كما يتيح للشبكة الترميزية والتفكيكية التقاط المعلومات الدلالية المهمة للتفاصيل واستعادة الدقة المكانية بالترتيب. ويستخدم نظام الاتصالات التجاوزية لتحسين تجزئة الصورة من خلال الاستفادة من المعلومات الدلالية العالية المستوى التي تم استخراجها في المسار الترميزي.

الفصل 4: مجموعات البيانات Datasets

مجموعة البيانات هي مجموعة من البيانات المخزنة بطريقة منظمة. وفي سياق التعلم الآلي، تعتبر مجموعة البيانات مجموعة من الأمثلة التي يتم استخدامها لتدريب أو تقييم نموذج التعلم الآلي. تتكون مجموعة البيانات عادةً من مجموعة من المتغيرات أو الميزات المدخلة ومجموعة مقابلة من المتغيرات الناتجة أو التسميات. تمثل الميزات البيانات المدخلة التي يتم استخدامها لإجراء التنبؤات، بينما تمثل التسميات البيانات الناتجة التي يتم تدريب نموذج التعلم الآلي للتنبؤ بها، حيث أن المهمة النهائية لنماذج التعلم الآلي هي الربط بين مجموعة مدخلات X ومجموعة مخرجات Y عن طريق إيجاد تابع تقريبي $F(x) = y$ ، وهنا تكون مجموعة البيانات هي عبارة عن الثنائيات (x, y) -وذلك في حالة التعلم بإشراف- التي تمثل المهمة التي سيتعلمها نموذج تعلم الآلة.



How data fits into Machine Learning Projects31 figure

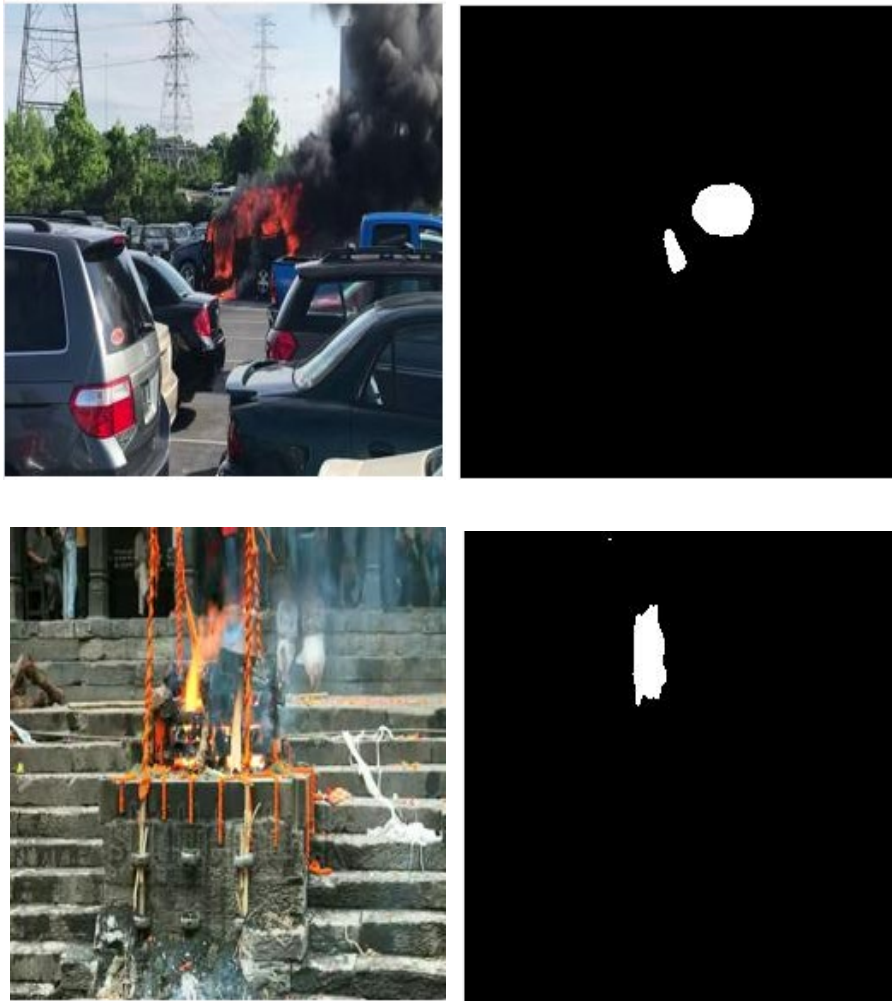
تُعد مجموعات البيانات عنصراً حاسماً في أي مشروع تعلم آلي، حيث تشكل الأساس الذي يبنى عليه نماذج التعلم الآلي وتُقيم عليها. تُعد جودة وكمية البيانات المستخدمة في تدريب النموذج عوامل حاسمة تحدد دقة وقدرات التعميم للنموذج الناتج. ولذلك، فإن اختيار مجموعة بيانات مناسبة والقيام بمعالجة البيانات الأولية بشكل كافٍ يعتبر خطوات حاسمة في أي مشروع تعلم آلي.

تختلف مصادر مجموعات البيانات التي تستخدم في مشاريع التعلم الآلي، حيث يمكن أن تكون من مصادر متعددة، مثل مستودعات البيانات العامة (Public Data Centres)، أرشيفات البيانات على الإنترنت، أو حتى البيانات التي يتم جمعها من فريق التعلم الآلي نفسه. وبمجرد الحصول على مجموعة بيانات، يتم إجراء مرحلة معالجة البيانات الأولية، والتي تتضمن تنظيف وتحويل البيانات وتحضيرها للاستخدام في نموذج التعلم الآلي. وتشمل هذه العملية غالباً إزالة البيانات غير الصالحة أو غير المكتملة، وتطبيع البيانات، وتحويلها إلى تنسيق مناسب يمكن تغذيته إلى خوارزمية التعلم الآلي.

يعد تفصيل حدود الحريق مهمة صعبة في الرؤية الحاسوبية، حيث تتضمن تحديد مناطق في صورة أو إطار فيديو تحتوي على النيران أو الدخان. يعتبر تفصيل الحريق بدقة أمراً حاسماً في العديد من التطبيقات، بما في ذلك الإطفاء، والاستجابة للكوارث، والسلامة الصناعية. ولتطوير نماذج التعلم الآلي التي يمكنها تفصيل الحريق بدقة في الصور ومقاطع الفيديو، يلزم توفر مجموعة بيانات عالية الجودة لصور الحرائق المصنفة. وفي هذا القسم، سنناقش مجموعتي بيانات تفصيل الحرائق الشائعتين

1. Fire Segmentation Image Dataset (FSID) (Kaggle)

هي مجموعة بيانات مشهورة للصور المصنفة التي تستخدم في مهام تفصيل الحرائق. تتألف المجموعة من أكثر من 20000 صورة تم تقسيمها إلى 3 فئات: صور تحتوي على نيران، وصور لا تحتوي على نيران، وصور ذات محتوى غامض. ويتم تصنيف كل صورة بقناع يشير إلى موقع الحريق في الصورة. المجموعة متاحة على Kaggle، وهي منصة عبر الإنترنت مشهورة لعلم البيانات ومحبي التعلم الآلي. تعد مجموعة بيانات تفصيل الحرائق 1 مصدراً مفيداً لتطوير وتقييم نماذج تفصيل الحرائق، وقد تم استخدامها في العديد من الدراسات البحثية. [28]



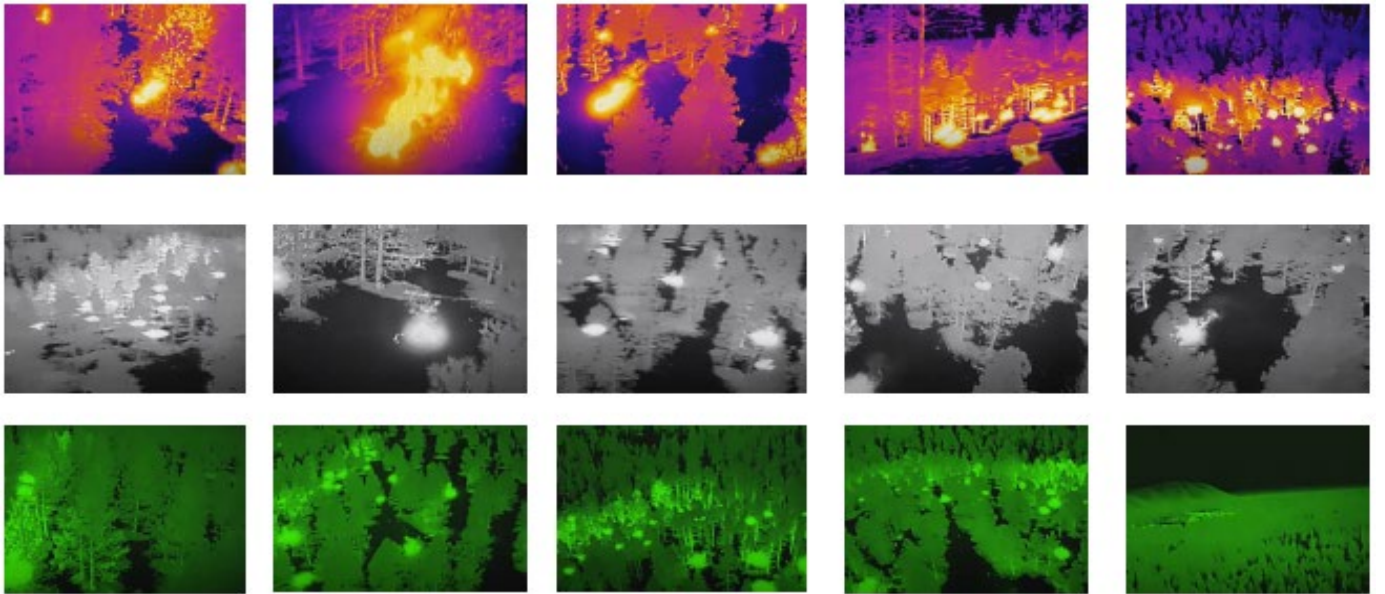
Images Sample from FSID dataset (left is the original RGB 32 figure image and right is the corresponding segmentation mask)

2. مجموعة البيانات FLAME:

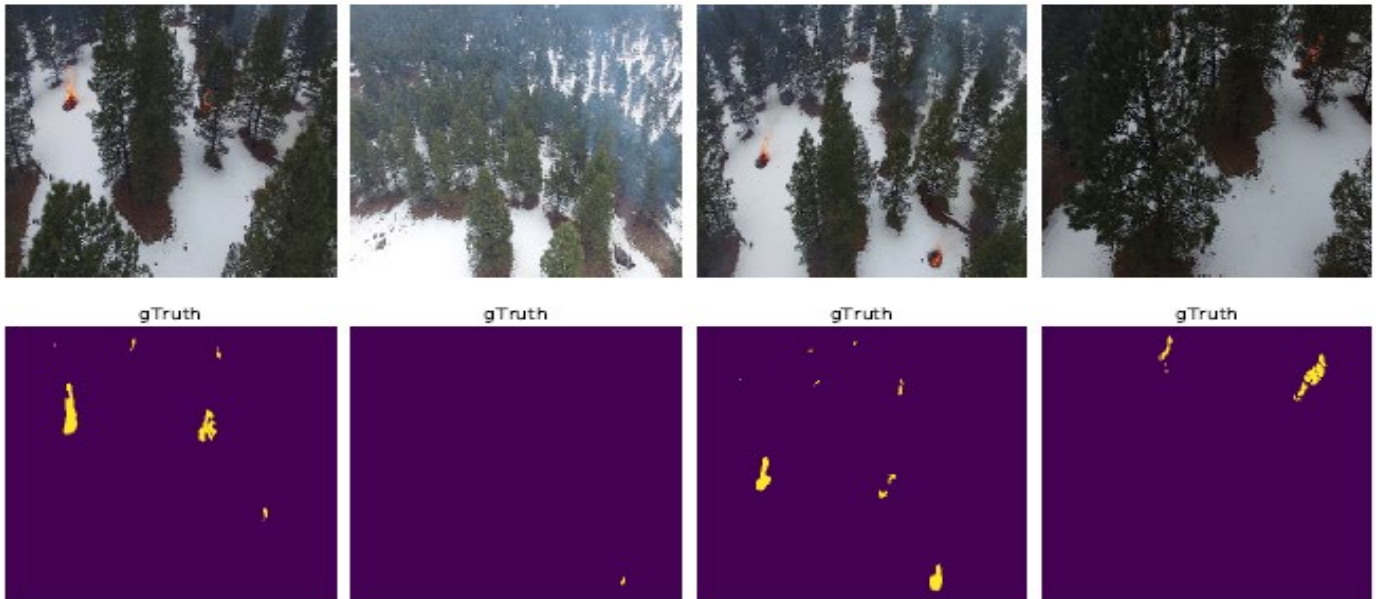
هي مجموعة بيانات شائعة أخرى تستخدم لمهام تفصيل الحرائق. تتألف المجموعة من صور جوية عالية الدقة التي تم التقاطها بواسطة مركبات جوية بدون طيار أو طائرات بدون طيار. تم التقاط الصور في بيئات مختلفة، بما في ذلك الغابات والمروج. تحتوي المجموعة على أكثر من 2000 صورة، وتم تصنيف كل صورة بقناع يشير إلى موقع الحريق في الصورة. تعد مجموعة بيانات FLAME تحدياً لتفصيل الحرائق، حيث تحتوي الصور على خلفيات معقدة ومناطق الحريق يمكن أن تكون صغيرة وغير منتظمة الشكل. المجموعة متاحة عبر الإنترنت وقد تم استخدامها في العديد من الدراسات البحثية لتطوير وتقييم نماذج تفصيل الحرائق.[29]



.Frame samples of the normal spectrum palette33 figure



Frame samples of thermal images including Fusion, WhiteHot, and GreenHot34 figure



Segmentation Sample35 figure

3. مجموعة بيانات KITTI:

KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) هي واحدة من أكثر مجموعات البيانات شهرة للاستخدام في الروبوتات المتحركة والقيادة الذاتية. يتكون من ساعات من سيناريوهات المرور المسجلة بمجموعة متنوعة من أنماط المستشعرات، بما في ذلك RGB عالية الدقة وكاميرات ستيريو رمادية وماسح ضوئي ثلاثي الأبعاد. يحتوي المعيار الكامل على العديد من المهام مثل السنيريو والتدفق البصري والقياس البصري والكشف عن الأشياء ثلاثية الأبعاد والتتبع ثلاثية الأبعاد. تحتوي مكتبة TensorFlow Datasets على مجموعة بيانات تحتوي على مجموعة بيانات كشف الأشياء، بما في ذلك الصور المنفردة ومربعات التحديد.[30]

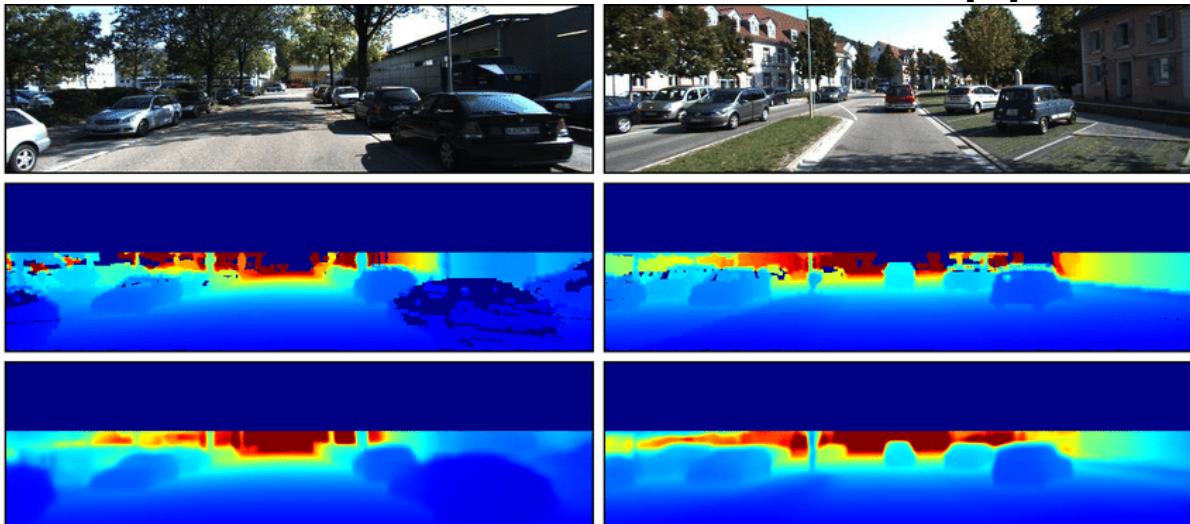


Image Sample from KITTI Dataset36 figure

وتحتوي على أكثر من 12000 مشهد خارجي تم التقاطها من سيارة متحركة. تم التقاط الصور في ظروف جوية مختلفة، بما في ذلك ضوء الشمس الساطع والأحوال الغائمة والمطر.

يتم تسمية كل صورة في مجموعة البيانات بخرائط عمق و تباين استريو. توفر خرائط عمق الحقيقية المسافات الفعلية بين الكاميرا والأشياء في المشهد، في حين توفر خرائط التباين استريو قياس الفرق في مواقع البكسلات بين صور الكاميرا اليسرى واليمنى. تعتبر هذه التسميات حاسمة لتدريب نماذج التعلم الآلي لمهام تقدير العمق.

KITTI هي مجموعة بيانات مستخدمة على نطاق واسع في مهام تقدير العمق، حيث تتضمن المشاهد بيانات معقدة بما في ذلك شوارع المدينة والطرق السريعة والطرق الريفية. كما تتضمن المجموعة بعض العوائق مثل المركبات والمشاة والدراجات، مما يمكن أن يشكل تحدياً لتقدير العمق بدقة. ومع ذلك، فإن مجموعة البيانات تُستخدم على نطاق واسع في مجتمع البحث في رؤية الحاسوب وقد تم استخدامها لتطوير نماذج تقدير العمق المتطورة.

4. مجموعة بيانات *NYU Depth*:

تعد مجموعة بيانات العمق الخاصة بجامعة نيويورك الإصدار 2 مجموعة بيانات مستخدمة على نطاق واسع من الصور الموصوفة المستخدمة لمهام تقدير العمق. تحتوي مجموعة البيانات على أكثر من 1449 مشهداً داخلياً تم التقاطها بكاميرا Microsoft Kinect. تم التقاط الصور لبيئات داخلية غير مبالية، بما في ذلك غرف المعيشة وغرف النوم والمكاتب والمطابخ. تنقسم مجموعة البيانات إلى جزأين: التدريب والاختبار، مع 795 صورة تدريبية و 654 صورة اختبار. [31]



NYU Depth Dataset Sample37 figure

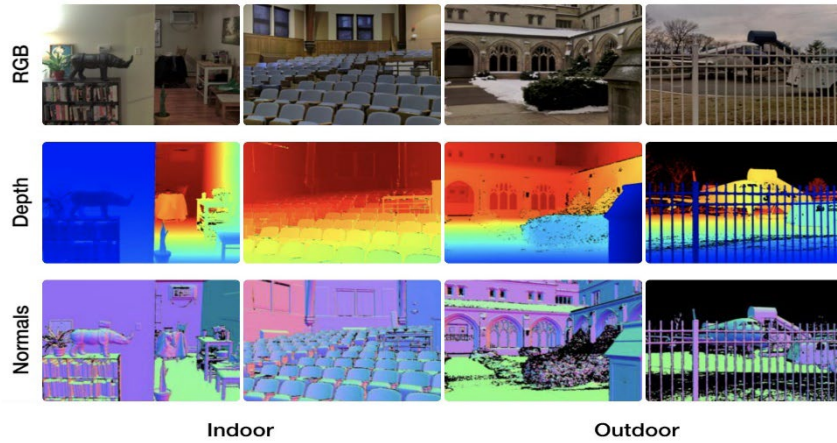
تتم تسمية كل صورة في مجموعة البيانات بخرائط عمق الحقيقة الأرضية، حيث يتم توفير المسافات الفعلية من الكاميرا إلى الكائنات في المشهد. تتضمن مجموعة البيانات أيضاً صور RGB وصور مجزأة، والتي يمكن استخدامها لتقييم وتطوير الخوارزميات لتقدير العمق.

تعد مجموعة بيانات NYU Depth Dataset v2 مجموعة بيانات صعبة لتقدير العمق، حيث تحتوي المشاهد على بيانات داخلية معقدة مع عناصر مختلفة مثل الأثاث والجدران والأبواب والنوافذ. تتضمن مجموعة البيانات أيضاً موضوعات بشرية في بعض المشاهد، والتي يمكن أن تشكل تحديات في تقدير العمق بدقة. ومع ذلك، تُستخدم مجموعة البيانات على نطاق واسع في مجتمع أبحاث الرؤية الحاسوبية، وقد تم استخدامها لتطوير نماذج تقدير العمق الحديثة.

باختصار، فإن NYU Depth Dataset v2 عبارة عن مجموعة بيانات مستخدمة على نطاق واسع من الصور المصنفة المستخدمة في مهام تقدير العمق. تحتوي مجموعة البيانات على أكثر من 1449 مشهداً داخلياً تم التقاطها بكاميرا Microsoft Kinect وهي مقسمة إلى مجموعات تدريب واختبار. تتم تسمية كل صورة بخرائط عمق الحقيقة الأرضية، وتتضمن مجموعة البيانات أيضاً صور RGB وصور مجزأة. في حين أن الإصدار 2 من مجموعة بيانات العمق من جامعة نيويورك يمثل مجموعة بيانات صعبة، فهو مورد لا يقدر بثمن لتطوير وتقييم نماذج التعلم الآلي لمهام تقدير العمق في البيئات الداخلية.

5. مجموعة بيانات DOIDE :

مجموعة بيانات DIODE هي مجموعة بيانات واسعة النطاق من المشاهد الخارجية المصممة لمهام تقدير العمق. تتكون مجموعة البيانات من 37000 إطار RGB-D ، بما في ذلك 22000 إطار تدريب و 15000 إطار اختبار. تم جمع مجموعة البيانات باستخدام مركبة مصممة خصيصاً ومجهزة بكاميرا Velodyne HDL-64E LiDAR وكاميرا Point Gray Grasshopper2 تم جمع البيانات في بيئات خارجية متنوعة، مثل الشوارع الحضرية والطرق الريفية والطرق السريعة.[32]



Sample of Images from DOIDE dataset38 figure

يتم تمييز كل إطار في مجموعة بيانات DIODE بخرائط عمق الحقيقة الأرضية، مما يوفر معلومات عمق دقيقة لكل بكسل. تتضمن مجموعة البيانات أيضاً تسميات التجزئة الدلالية للأشياء الموجودة في المشهد، مثل السيارات والمباني والمشاة. تنقسم مجموعة البيانات أيضاً إلى مجموعات التدريب والتحقق والاختبار، مما يمكن الباحثين من تقييم أداء نماذجهم بطريقة مسيطر عليها وموحدة.

تشكل مجموعة بيانات DIODE تحديات فريدة لنماذج تقدير العمق، مثل التعامل مع البيانات الخارجية الكبيرة، والأشياء المعقدة، وظروف الإضاءة المتنوعة. ومع ذلك، تعد مجموعة بيانات DIODE مورداً أساسياً لتطوير وتقييم نماذج تقدير العمق التي يمكن أن تؤدي أداءً قوياً في السيناريوهات الخارجية في العالم الحقيقي.

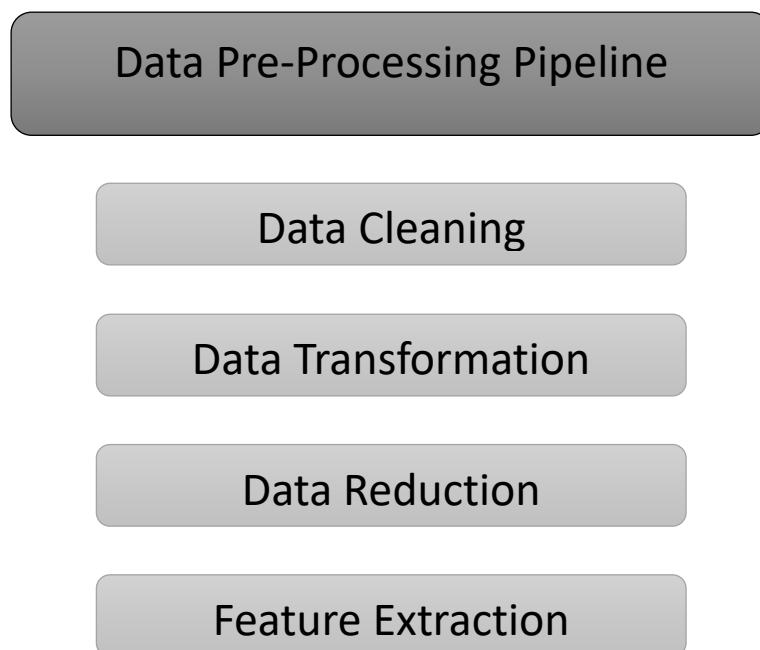
باختصار، مجموعة بيانات DIODE عبارة عن مجموعة بيانات واسعة النطاق من المشاهد الخارجية المصممة لمهام تقدير العمق. تتكون مجموعة البيانات من 37000 إطار RGB-D ، مُصنَّفة بخرائط عمق الحقيقة الأرضية وعلامات التجزئة الدلالية. تشكل مجموعة البيانات تحديات فريدة لنماذج تقدير العمق، مما يجعلها مورداً قيماً لتطوير وتقييم النماذج التي يمكن أن تؤدي أداءً قوياً في البيئات الخارجية.

في التعلم الآلي، تلعب جودة بيانات الإدخال دوراً مهماً في دقة ومتانة النموذج الناتج. غالباً ما تحتوي البيانات الأولية على ضوضاء وتناقضات وعيوب أخرى يمكن أن تؤثر سلباً على عملية التعلم. لذلك، تعد مرحلة ما قبل المعالجة خطوة حاسمة في إعداد البيانات لمهام التعلم الآلي. في هذا القسم، سنناقش أهمية المعالجة المسبقة للبيانات وبعض التقنيات الشائعة المستخدمة في هذه المرحلة.

4. مرحلة المعالجة المسبقة للبيانات

المعالجة المسبقة للبيانات هي عملية إعداد البيانات الأولية للتحليل عن طريق إزالة الضوضاء والتناقضات والعيوب الأخرى. تعد مرحلة ما قبل المعالجة ضرورية لمهام التعلم الآلي لأنها تؤثر بشكل مباشر على جودة بيانات الإدخال، وبالتالي على دقة ومتانة النماذج الناتجة. الهدف الرئيسي من المعالجة المسبقة للبيانات هو جعل البيانات أكثر ملاءمة لخوارزميات التعلم الآلي.

تتضمن مرحلة المعالجة المسبقة للبيانات عادةً الخطوات كما في المخطط التالي:



Data pre-processing pipeline39 figure

1. **تنظيف البيانات:** تتضمن هذه الخطوة إزالة أو تصحيح أي أخطاء أو تناقضات في البيانات. على سبيل المثال، إزالة التكرارات وتصحيح الأخطاء الإملائية ومعالجة القيم المفقودة.

2. **تحويل البيانات:** تتضمن هذه الخطوة تحويل البيانات إلى تنسيق مناسب للتحليل. على سبيل المثال، تحويل المتغيرات الفئوية إلى قيم عددية أو تحجيم القيم الرقمية إلى نطاق قياسي.

3. **تقليل البيانات:** تتضمن هذه الخطوة تقليل كمية البيانات عن طريق إزالة الميزات غير الملائمة أو الزائدة عن الحاجة. هذا يساعد على تبسيط النموذج وتحسين أدائه.

4. **استخراج الميزة:** تتضمن هذه الخطوة إنشاء ميزات جديدة من البيانات الموجودة. على سبيل المثال، استخراج ميزات النسيج من الصور أو اشتقاق متغيرات جديدة من المتغيرات الموجودة.

تتضمن الأساليب الشائعة المستخدمة في مرحلة المعالجة المسبقة للبيانات التطبيع، وقياس الميزات، وتقليل الأبعاد، واستخراج الميزات. يتضمن التطبيع تحجيم البيانات إلى نطاق مشترك، بينما يتضمن تحجيم الميزة تحجيم كل ميزة بحيث يكون لها متوسط صفري وتباين الوحدة. تعمل تقنيات تقليل الأبعاد، مثل تحليل المكونات الرئيسية (-PCA) Principle Component Analysis، على تقليل عدد الميزات من خلال إنشاء مجموعة جديدة من المتغيرات التي تلتقط أهم المعلومات في البيانات. تساعد تقنيات الكشف الخارجية في تحديد وإزالة أي نقاط بيانات تختلف اختلافاً كبيراً عن بقية البيانات.

في الختام، تعد المعالجة المسبقة للبيانات خطوة حاسمة في التعلم الآلي التي تساعد على تحسين جودة بيانات الإدخال، وبالتالي دقة ومتانة النماذج الناتجة. تتضمن مرحلة ما قبل المعالجة عدة خطوات، مثل تنظيف البيانات وتحويلها وتقليلها واستخراجها. تشمل التقنيات الشائعة المستخدمة في هذه المرحلة التطبيع، وقياس الميزات، وتقليل الأبعاد، والكشف عن الانحرافات.

الفصل 5: التجارب والنتائج

التجارب:

مهمة تجزئة النار (Fire Segmentation Task)

بالنسبة لمهمة تجزئة النار Fire Segmentation تم اختبار بنيتين مختلفتين: U-Net و DenseNet. تم تقييم أداء هذه البنى باستخدام معيار متوسط التقاطع إلى الاجتماع (Mean Intersection Over Union) (mIoU). كانت مجموعات البيانات المستخدمة لهذه المهمة هي مجموعة بيانات صور تجزئة النار (FSID) ومجموعة بيانات FLAME.

عند تدريب مثل لكل نموذج على كل مجموعة بيانات على حدة، وجد أن بنية U-Net تفوقت على بنية DenseNet في مجموعتي البيانات. ومع ذلك، عند التدريب على مجموعة بيانات FLAME ثم تحسين الضبط Fine-Tuning باستخدام مجموعة بيانات FSID، حققت بنية DenseNet نتائج أفضل إلى حد ما مما كانت عليه عندما تم تدريبها فقط على FSID. من المحتمل أن يكون هذا بسبب الاختلاف في التوزيع بين مجموعتي البيانات.

تقدير العمق (Depth Estimation Task)

بالنسبة لمهمة تقدير العمق، تم استخدام بنية U-Net ذات الصناديق التكيفية وحققت نتائج جيدة جداً على مقياس متوسط جذر الخطأ التربيعي (Root Mean Squared Error) مقارنة بمعمارية DenseNet المدربة على نفس البيانات. تم استخدام مجموعة بيانات DIODE للتدريب، مع إجراء الضبط الدقيق باستخدام مجموعات بيانات NYU Depth أو KITTI.

تم تحقيق أفضل النتائج عند التدريب على مجموعة بيانات DIODE والضبط الدقيق باستخدام مجموعة بيانات NYU Depth V2. تم تحقيق ثاني أفضل النتائج عند التدريب على DIODE والضبط الدقيق على KITTI. أدى التدريب على كل مجموعة بيانات على حدة إلى أداء ضعيف.

باختصار، لكلتا المهمتين، تفوقت بنية U-Net بشكل عام على بنية DenseNet. أدى الضبط الدقيق لمجموعة البيانات الثانية إلى تحسين الأداء لكلتا المهمتين. تم تحقيق أفضل النتائج عند الضبط الدقيق لمجموعة بيانات بتوزيع مشابه للبيانات المستهدفة.

المشاكل والحلول

يمكن أن يمثل العمل في مشروع التعلم العميق مثل تقسيم النار باستخدام تقدير العمق الأحادي عدداً من التحديات. سنناقش في هذا القسم بعض المشكلات التي تمت مواجهتها أثناء المشروع والحلول التي تم تنفيذها للتغلب عليها.

المعالجة المسبقة للبيانات (Data Pre-Processing):

كان أحد التحديات الأولى التي تمت مواجهتها يتعلق بالمعالجة المسبقة للبيانات. كانت مجموعة بيانات صور تجزئة النار (FSID) ومجموعة بيانات FLAME ، التي تم استخدامها لمهمة تجزئة النار ، ذات أحجام وتنسيقات مختلفة للصور. يتطلب هذا معالجة مسبقة دقيقة لضمان تغيير حجم الصور وتطبيعها بشكل صحيح قبل إدخالها في النموذج.

لحل هذه المشكلة، تم تطوير خط أنابيب مخصص للمعالجة المسبقة. تضمن خط الأنابيب هذا خطوات لتغيير حجم الصور إلى حجم شائع، وتسوية قيم البكسل الخاصة بها. يضمن ذلك أن الصور من مجموعتي البيانات كانت بتنسيق ثابت ويمكن استخدامها بالتبادل أثناء التدريب.

اختيار النموذج (Model Selection):

كان التحدي الآخر هو اختيار بنية النموذج المناسبة لكل مهمة. بالنسبة لمهمة Fire Segmentation تم اختبار بنيتين مختلفتين U-Net و DenseNet: تقييم أداء هذه البنى باستخدام المقياس المحدد للمهمة، من الجدير بالذكر أن هناك العديد من البنى الأخرى التي كانت محط الاهتمام للاستخدام في مهمة تجزئة النار كالشبكة الشهيرة YoloV5 والتي تستخدم غالباً في مهام كشف الأغراض، وشبكة Masked R-CNN والتي قدمت أفضل أداء في العديد من المهام، لكن العائق الأكبر كان في أن شبكة Yolo V5 تقوم بتصغير حجم البيانات المدخلة من بداية الانتشار إلى نهايته ولا يوجد قسم في الشبكة لترميم المعلومات (تكبير) بشكل تدريجي وصولاً إلى خرج بنفس حجم الدخل أي أنها ليست شبكة مصممة لهذا النوع من المسائل وإجراء التجارب سيتطلب كثيراً من الوقت والموارد الغير متاحة.

ضبط الهايبربارامترات (Hyperparameters Tuning):

يعد ضبط Hyperparameter خطوة مهمة في أي مشروع تعلم عميق. في هذا المشروع، تم ضبط العديد من المعلمات الفائقة لتحسين أداء النموذج، بما في ذلك معدل التعلم وحجم الدفعة وعدد العصور.

تم إجراء بحث في الشبكة (Grid Search) للعثور على التركيبة المثلى للمعلمات الفائقة لكل مهمة. تضمن ذلك تدريب نماذج متعددة بمجموعات مختلفة من المعلمات (الهابيربارامترات) وتقييم أدائها على مجموعة التحقق.

تم العثور على المعلمات التشعبية المثلى لتكون مختلفة لكل مهمة ومجموعة بيانات. على سبيل المثال، وجد أن معدل التعلم المنخفض $\alpha = 0.0001$ يكون أكثر فعالية عند التدريب على مجموعة بيانات FSID مقارنة بمجموعة بيانات (Flame) حيث وجد معامل التعلم الأفضل $\alpha = 0.001$ وبالمثل، وجد أن حجم الدفعة الأكبر (32) يكون أكثر فاعلية عند التدريب على مجموعة بيانات DIODE مقارنة بمجموعات بيانات NYU Depth أو KITTI.

إفراط التعلم (Overfitting):

يعد فرط التعلم مشكلة شائعة في مشاريع التعلم العميق. يحدث عندما يحفظ النموذج بيانات التدريب بدلاً من تعلم التعميم على البيانات الجديدة. يمكن أن يؤدي هذا إلى ضعف الأداء عند تقييم النموذج في مجموعة اختبار. لمنع فرط التخصيص، تم استخدام العديد من التقنيات أثناء التدريب. وشمل ذلك التوقف المبكر وزيادة البيانات وتنظيمها. يتضمن التوقف المبكر مراقبة أداء النموذج على مجموعة التحقق أثناء التدريب وإيقاف التدريب مبكراً إذا بدأ الأداء في التدهور. هذا يمكن أن يمنع فرط التجهيز عن طريق منع النموذج من حفظ بيانات التدريب.

تتضمن زيادة البيانات زيادة حجم مجموعة التدريب بشكل مصطنع عن طريق تطبيق تحويلات عشوائية على الصور. يمكن أن يساعد هذا في منع فرط التخصيص من خلال توفير المزيد من البيانات المتنوعة للنموذج للتعلم منها.

يتضمن التنظيم إضافة مصطلح جزائي إلى وظيفة الخسارة التي تشجع النموذج على الحصول على أوزان صغيرة. يمكن أن يساعد هذا في منع فرط التخصيص عن طريق تثبيط النموذج عن الاعتماد بشدة على أي ميزة واحدة. باستخدام هذه التقنيات، تم منع فرط التخصيص بنجاح وتم تحقيق أداء التعميم الجيد في كلتا المهمتين.

النتائج

عند تدريب مثيل لكل نموذج على كل مجموعة بيانات على حدة، وجد أن بنية U-Net تفوقت على بنية DenseNet في مجموعتي البيانات. ومع ذلك، عند التدريب على مجموعة بيانات FLAME ثم ضبط مجموعة بيانات FSID، حققت بنية DenseNet نتائج أفضل إلى حد ما مما كانت عليه عندما تم تدريبها فقط على FSID. من المحتمل أن يكون هذا بسبب الاختلاف في التوزيع بين مجموعتي البيانات.

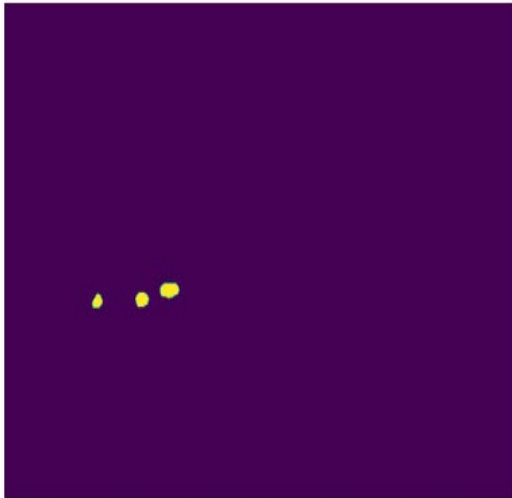
Task	Architecture	Dataset	Fine-tuning	Performance
Fire Segmentation	U-Net	FSID	N/A	0.8
Fire Segmentation	DenseNet	FSID	N/A	0.73
Fire Segmentation	U-Net	FLAME	N/A	0.7
Fire Segmentation	DenseNet	FLAME	N/A	0.69
Fire Segmentation	DenseNet	FLAME	FSID	0.74
Depth Estimation	Adaptive Bins U-Net	DIODE	NYU Depth	0.09
Depth Estimation	Adaptive Bins U-Net	DIODE	KITTI	0.11
Depth Estimation	Adaptive Bins U-Net	DIODE	N/A	0.3
Depth Estimation	DenseNet	DIODE	N/A	0.28

تم تحقيق أفضل النتائج لتقدير العمق عند التدريب على مجموعة بيانات DIODE والضبط الدقيق باستخدام عمق جامعة نيويورك. تم تحقيق ثاني أفضل النتائج عند التدريب على DIODE والضبط الدقيق على KITTI. أدى التدريب على كل مجموعة بيانات على حدة إلى أداء ضعيف.

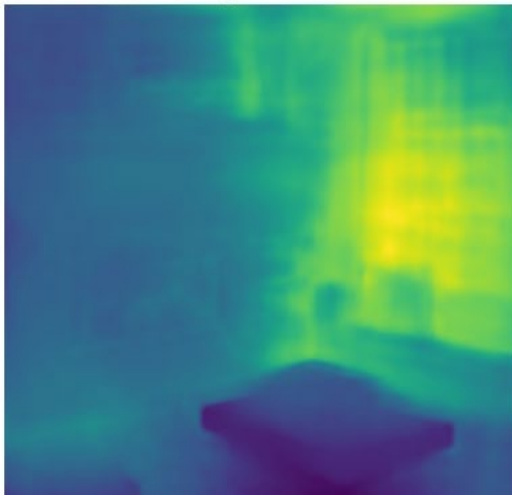
Sample Image 1



Fire Mask



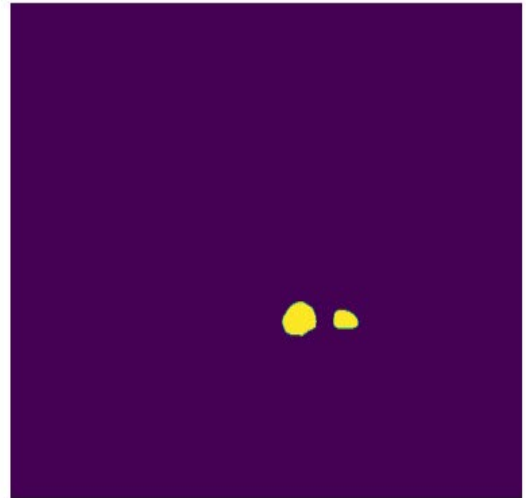
Depth Mask



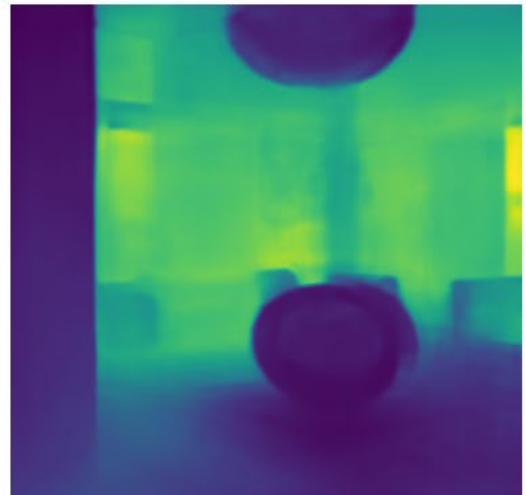
Sample Image 2



Fire Mask



Depth Mask



سلبيات النظام:

بينما أظهر نظامنا لتجزئة النار بتقدير العمق الأحادي نتائج واعدة، إلا أنه لا يخلو من القيود. أحد القيود الرئيسية هو أن النظام غير مبني على مجموعة بيانات مصممة خصيصاً للعثور على بعد النار بالنسبة للكاميرا. بدلاً من ذلك، تم تصميمه كمزيج من نموذجين: أحدهما لتقسيم الصور إلى بكسلات نارية وغير نارية، والآخر للتنبؤ بعمق كل بكسل دون مراعاة الكائن الذي يمثل هذا البكسل.

هذا النهج له عدة عيوب. أولاً، لم تحتوي أي من مجموعات بيانات العمق المستخدمة في تجاربنا على أي مشاهد للنار. هذا يعني أن نموذج تقدير العمق الخاص بنا لم يرَ حريقاً مطلقاً، وبالتالي فهو غير قادر على استنتاج المسافة الدقيقة للحريق بالنسبة للصورة بدقة. بدلاً من ذلك، يستنتج النظام مسافة النار فيما يتعلق بالأشياء القريبة ثم يقيسها. يمكن أن يؤدي هذا إلى نتائج غير دقيقة وقد لا يوفر دائماً تقديراً موثقاً لمسافة الحريق.

يتمثل أحد القيود الأخرى لنظامنا في أنه يعتمد على نموذجين منفصلين لتجزئة النار وتقدير العمق. هذا يعني أنه لا يوجد تفاعل مباشر بين هاتين المهمتين ولا توجد طريقة للنظام للاستفادة من المعلومات من مهمة واحدة لتحسين الأداء على الأخرى. على سبيل المثال، إذا حدد نموذج تجزئة الحريق منطقة تحتوي على حريق، فمن المحتمل أن يتم استخدام هذه المعلومات بواسطة نموذج تقدير العمق لتحسين تنبؤاته في تلك المنطقة. ومع ذلك، نظراً لأن نظامنا يستخدم نموذجين منفصلين، فإن هذا النوع من التفاعل غير ممكن.

علاوة على ذلك، لا يأخذ نظامنا في الاعتبار أي معلومات سياقية عند إجراء تنبؤاته. على سبيل المثال، لا يأخذ في الاعتبار عوامل مثل ظروف الإضاءة أو وجود دخان أو عوائق بصرية أخرى عند تقدير العمق. يمكن أن يؤدي هذا إلى نتائج غير دقيقة في السيناريوهات الصعبة حيث تلعب هذه العوامل دوراً مهماً.

باختصار، في حين أن نظامنا لتجزئة الحرائق بتقدير العمق الأحادي قد أظهر نتائج واعدة، فإنه لا يخلو من قيودها. يمكن أن يركز العمل المستقبلي على معالجة هذه القيود من خلال تطوير مجموعة بيانات مصممة خصيصاً لهذه المهمة، ودمج تجزئة الحريق وتقدير العمق في نموذج واحد، ودمج المعلومات السياقية في عملية التنبؤ.

ملخص:

في هذا التقرير، استكشفنا استخدام تقنيات الذكاء الاصطناعي، وتحديدًا التعلم العميق والشبكات العصبية التلافيفية، لمهام تجزئة النار وتقدير العمق الأحادي. لقد ناقشنا أهمية اكتشاف الحرائق وتحديد مواقعها للحفاظ على السلامة العامة وكيف يمكن لتقنيات الذكاء الاصطناعي مثل شبكة الأمم المتحدة أن تلعب دوراً مهماً في تحقيق هذا الهدف.

لقد ناقشنا أيضاً العديد من التقنيات المستخدمة في التقسيم الدلالي وتقدير العمق، بما في ذلك U-Net و SegNet و Masked RCNN و U-Net المتكيف. بالإضافة إلى ذلك، استكشفنا استخدام مجموعات البيانات مثل مجموعة بيانات صورة تقسيم النار FSID ومجموعة بيانات FLAME ومجموعة بيانات DIODE ومجموعة بيانات NYU Depth V2 ومجموعة بيانات KITTI للتدريب وتقييم نماذج التعلم الآلي لهذه المهام.

أظهرت تجاربنا نتائج واعدة لكل من مهام تجزئة النار وتقدير العمق الأحادي باستخدام نماذج التعلم العميق. ومع ذلك، فإن نظامنا لا يخلو من قيودها. أحد القيود الرئيسية هو أن نظامنا غير مبني على مجموعة بيانات مصممة خصيصاً للعثور على مسافة النار بالنسبة للكاميرا. بدلاً من ذلك، تم تصميمه كمزيج من نموذجين: أحدهما لتقسيم الصور إلى وحدات بكسل نارية وغير نارية، والآخر للتنبؤ بعمق كل بكسل بغض النظر عن الكائن الذي يمثلته هذا البكسل.

يمكن أن يركز العمل المستقبلي على معالجة هذه القيود من خلال تطوير مجموعة بيانات مصممة خصيصاً لهذه المهمة، ودمج تجزئة الحريق وتقدير العمق في نموذج واحد، ودمج المعلومات السياقية في عملية التنبؤ. بشكل عام، أظهر نظامنا نتائج واعدة ولديه القدرة على لعب دور مهم في تحسين السلامة العامة من خلال الكشف الدقيق عن الحرائق وتحديد مواقعها.

الأعمال المستقبلية

في المستقبل، هناك العديد من مجالات البحث والتطوير المحتملة التي يمكن أن تبني على العمل المقدم في هذا التقرير. وتشمل هذه:

- 1- تطوير مجموعة بيانات مصممة خصيصاً لإيجاد مسافة النار بالنسبة للكاميرا: أحد القيود الرئيسية لنظامنا الحالي هو أنه غير مبني على مجموعة بيانات مصممة خصيصاً لهذه المهمة. يمكن أن يؤدي تطوير مجموعة البيانات هذه إلى تحسين دقة نظامنا وموثوقيته بشكل كبير.

- 2- دمج تجزئة الحريق وتقدير العمق في نموذج واحد: يستخدم نظامنا الحالي نموذجين منفصلين لتجزئة النار وتقدير العمق. قد يؤدي دمج هاتين المهمتين في نموذج واحد إلى السماح بالتفاعل المباشر بين المهمتين وتمكين النظام من الاستفادة من المعلومات من مهمة واحدة لتحسين الأداء على الأخرى.
- 3- دمج المعلومات السياقية في عملية التنبؤ: لا يأخذ نظامنا الحالي في الاعتبار أي معلومات سياقية عند إجراء تنبؤاته. قد يؤدي دمج عوامل مثل ظروف الإضاءة أو وجود دخان أو عوائق بصرية أخرى في عملية التنبؤ إلى تحسين دقة نظامنا في السيناريوهات الصعبة.
- 4- استكشاف بنى التعلم العميق البديلة: بينما أظهرت تجاربنا نتائج واعدة باستخدام معماريات U-Net و DenseNet، فقد تكون هناك بنى تعليمية عميقة أخرى يمكنها تحقيق أداء أفضل في هذه المهام.
- 5- تصميم روبوت للقيام بمهمة اكتشاف النار باستخدام هذه النماذج، وإطفاؤها باستخدام التقنيات الحديثة، حيث يمكن تحسين النتائج ورفع دقتها باستخدام حساسات خاصة يتم تفعيلها عند اقتراب الروبوت من الموقع التقديري للنار.

بشكل عام، هناك العديد من الفرص المثيرة للبحث والتطوير في المستقبل في مجال تجزئة النار القائمة على الذكاء الاصطناعي وتقدير العمق الأحادي. من خلال البناء على العمل المقدم في هذا التقرير، يمكننا الاستمرار في تحسين دقة وموثوقية هذه الأنظمة وإحراز تقدم كبير نحو تحسين السلامة العامة.

مجالات التحسين

بالإضافة إلى الأعمال المستقبلية المذكورة في القسم السابق، هناك العديد من المجالات التي يمكن تحسين نظامنا الحالي فيها. وتشمل هذه:

- 1- تحسين المعالجة المسبقة للبيانات: تعد المعالجة المسبقة للبيانات خطوة حاسمة في التعلم الآلي تساعد على تحسين جودة بيانات الإدخال. يمكن أن يستفيد نظامنا الحالي من المزيد من التحسينات في تقنيات المعالجة المسبقة للبيانات مثل تنظيف البيانات، وتحويل البيانات، والتطبيع، وتوسيع نطاق الميزات، وتقليل الأبعاد، واكتشاف الشذوذ.
- 2- الصقل الدقيق لاختيار النموذج وضبط المعلمة الفائقة: يمكن أن يكون لاختيار بنية النموذج المناسبة وضبط المعلمة الفائقة تأثير كبير على أداء نماذج التعلم الآلي. يمكن أن يساعد إجراء المزيد من التجارب مع بنى النماذج المختلفة وإعدادات المعامل الفائقة في تحسين أداء نظامنا.
- 3- معالجة فرط التخصيص: يعد Overfitting مشكلة شائعة في التعلم الآلي حيث يعمل النموذج جيداً على بيانات التدريب ولكن بشكل سيئ على البيانات الجديدة. يمكن استخدام تقنيات مثل التنظيم والتوقف المبكر وزيادة البيانات للمساعدة في منع التجهيز الزائد وتحسين قدرة التعميم لنماذجنا.

4- دمج معلومات إضافية في عملية التنبؤ: يعتمد نظامنا الحالي فقط على المعلومات المرئية من الصور لعمل تنبؤاته. يمكن أن يساعد دمج معلومات إضافية مثل قراءات درجة الحرارة أو أجهزة إنذار كاشفات الدخان في عملية التنبؤ في تحسين دقة نظامنا.

بشكل عام، هناك العديد من المجالات التي يمكن تحسين نظامنا الحالي فيها. من خلال معالجة مجالات التحسين هذه والبناء على العمل المقدم في هذا التقرير، يمكننا الاستمرار في إحراز تقدم نحو تطوير أنظمة أكثر دقة وموثوقية قائمة على الذكاء الاصطناعي لتجزئة الحرائق وتقدير العمق الأحادي.

References:

- 1- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9, 381-386.
- 2- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*. b, 4, 51-62.
- 3- Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2018, October). Unsupervised learning based on artificial neural network: A review. In *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)* (pp. 322-327). IEEE.
- 4- Levine, S., Kumar, A., Tucker, G., & Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*.
- 5- Zou, X., Hu, Y., Tian, Z., & Shen, K. (2019, October). Logistic regression model optimization and case analysis. In *2019 IEEE 7th international conference on computer science and network technology (ICCSNT)* (pp. 135-139). IEEE.
- 6- Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39, 261-283.
- 7- Bhavsar, H., & Panchal, M. H. (2012). A review on support vector machine for data classification. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(10), 185-189.
- 8- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015). <https://doi.org/10.1038/nature14539>.
- 9- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115-133.
- 10- Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. *Backpropagation: Theory, architectures and applications*, 1-34.
- 11- Thoma, M. (2016). A survey of semantic segmentation. *arXiv preprint arXiv:1602.06541*.
- 12- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- 13- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18 (pp. 234-241). Springer International Publishing.
- 14- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- 15- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
- 16- Eigen, D., Puhrsch, C., & Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27.
- 17- Zhou, W., Zhou, E., Liu, G., Lin, L., & Lumsdaine, A. (2019). Unsupervised monocular depth estimation from light field image. *IEEE Transactions on Image Processing*, 29, 1606-1617.
- 18- Xu, Y., Yang, X., Yu, Y., Jia, W., Chu, Z., & Guo, Y. (2022). Depth Estimation by Combining Binocular Stereo and Monocular Structured-Light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1746-1755).
- 19- Xu, Y., Yang, X., Yu, Y., Jia, W., Chu, Z., & Guo, Y. (2022). Depth Estimation by Combining Binocular Stereo and Monocular Structured-Light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1746-1755).

- 20- Chen, S. J., Hovde, D. C., Peterson, K. A., & Marshall, A. W. (2007). Fire detection using smoke and gas sensors. *Fire Safety Journal*, 42(8), 507-515.
- 21- Tadic, V., Toth, A., Vizvari, Z., Klincsik, M., Sari, Z., Sarcevic, P., ... & Biro, I. (2022). Perspectives of Realsense and ZED depth sensors for robotic vision applications. *Machines*, 10(3), 183.
- 22- Zhu, S. C., & Yang, M. H. (2014). Object recognition: From human vision to machine vision. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(2), 1-197.
- 23- Toderici, G., Vincent, D., Johnston, N., Hwang, Y., Minnen, D., Covell, M., ... & Shor, J. (2017). Full resolution image compression with recurrent neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5306-5314).
- 24- Wallace, G. K. (1992). The JPEG still picture compression standard. *IEEE transactions on consumer electronics*, 38(1), xviii-xxxiv.
- 25- Boutell, T., et al. (1999). PNG (Portable Network Graphics) specification, version 1.2. W3C Recommendation, World Wide Web Consortium.
- 26- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354-377.
- 27- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- 28- Zhang, Q., Xu, J., Xu, L., & Guo, H. (2016, January). Deep convolutional neural networks for forest fire detection. In *2016 International Forum on Management, Education and Information Technology Application* (pp. 568-575). Atlantis Press.
- 29- Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P. Z., & Blasch, E. (2021). Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Computer Networks*, 193, 108001.
- 30- Deschaud, J. E. (2021). KITTI-CARLA: a KITTI-like dataset generated by CARLA Simulator. *arXiv preprint arXiv:2109.00892*.
- 31- Couprie, C., Farabet, C., Najman, L., & LeCun, Y. (2013). Indoor semantic segmentation using depth information. *arXiv preprint arXiv:1301.3572*.
- 32- Vasiljevic, I., Kolkin, N., Zhang, S., Luo, R., Wang, H., Dai, F. Z., ... & Shakhnarovich, G. (2019). Diode: A dense indoor and outdoor depth dataset. *arXiv preprint arXiv:1908.00463*.