**الجامعة السورية الخاصة**



**بناء أداة لتحسين المساحات الداخلية(جدران/أرضيات)**

**إعداد الطالبات:**

**أية سحلول**

**جودي المصري**

**أية الطويل**

**بإشراف الدكتورة:**

**ماجدة البكور**

**المهندسة وسام السحلي**

**الفهرس:**

جدول المحتويات

[**الفصل الأول: المقدمة** 7](#_Toc220538993)

[1.1 خلفية البحث: 7](#_Toc220538994)

[1.2 مشكلة البحث: 8](#_Toc220538995)

[1.3 مسائل البحث: 8](#_Toc220538996)

[1.4 أثر وأهمية البحث: 8](#_Toc220538997)

[**الفصل الثاني: الدراسة المرجعية للتطبيقات والأبحاث ذات الصلة بالمشروع:** 9](#_Toc220538998)

[2.1 الدراسة المرجعية للتطبيقات والبرامج المشابهة: 9](#_Toc220538999)

[2.2 الدراسة المرجعية للأبحاث والأوراق العلمية: 12](#_Toc220539000)

[**الفصل الثالث: الدراسة النظرية** 15](#_Toc220539001)

[3.1 مقدّمة: 15](#_Toc220539004)

[3.2 أساسيات التعلّم العميق: 15](#_Toc220539005)

[**3.2.1** **مفهوم التعلّم العميق ودوافع استخدامه:** 15](#_Toc220539010)

[**3.2.2** **آلية التعلّم داخل الشبكات العميقة:** 16](#_Toc220539016)

[3.3 الشبكات العصبية الالتفافية (CNNs): 16](#_Toc220539017)

[**3.3.1** **بنية الشبكات الالتفافية وأهميتها:** 16](#_Toc220539023)

[**3.3.2** **دور الشبكات الالتفافية في فهم المشاهد الداخلية:** 16](#_Toc220539030)

[3.4 نماذج Transformers في الرؤية الحاسوبية: 17](#_Toc220539031)

[**3.4.1** **التحول نحو النماذج السياقية:** 17](#_Toc220539038)

[**3.4.2** **آلية الانتباه الذاتي ودورها:** 17](#_Toc220539046)

[3.5 التقسيم الدلالي للصور: 18](#_Toc220539047)

[**3.5.1** **مفهوم التقسيم الدلالي:** 18](#_Toc220539055)

[**3.5.2** **أهمية التقسيم الدلالي في المشروع:** 18](#_Toc220539064)

[3.6 نموذج :Segment Anything Model (SAM) 19](#_Toc220539065)

[**3.6.1** **لمحة عامة عن النموذج:** 19](#_Toc220539074)

[**3.6.2** **دوره في تحسين الدقة الهندسية:** 19](#_Toc220539084)

[3.7 تكامل المفاهيم ضمن النظام المقترح: 19](#_Toc220539085)

[3.8 خلاصة الفصل: 20](#_Toc220539086)

[**الفصل الرابع: المعطيات الأولية وطرق تحصيلها** 20](#_Toc220539087)

[4.1 مقدّمة: 21](#_Toc220539091)

[4.2 أنواع المعطيات الأولية المعتمدة في المشروع: 21](#_Toc220539094)

[4.3 المعطيات الأولية الأولى: صور مشاهد داخلية موسومة: 21](#_Toc220539095)

[4.4 المعطيات الأولية الثانية: صور مشاهد داخلية واقعية غير موسومة: 21](#_Toc220539096)

[4.5 طريقة تحصيل المعطيات: 22](#_Toc220539097)

[4.6 تنظيف المعطيات وتجهيزها: 22](#_Toc220539098)

[4.7 استخدام المعطيات ضمن النظام: 22](#_Toc220539099)

[4.8 خلاصة الفصل: 23](#_Toc220539100)

[**الفصل الخامس :المنهجية المقترحة** 23](#_Toc220539101)

[5.1 مقدّمة الفصل: 23](#_Toc220539103)

[5.2 بنية النظام العامة: 24](#_Toc220539104)

[5.3 نموذج تصنيف نوع الغرفة: 24](#_Toc220539105)

[**5.3.1** **دور النموذج ضمن النظام:** 25](#_Toc220539113)

[**5.3.2** **آلية عمل النموذج:** 25](#_Toc220539122)

[**5.3.3** **القيمة المضافة للنموذج:** 25](#_Toc220539132)

[5.4 نموذج SegFormer للتقسيم الدلالي: 26](#_Toc220539133)

[**5.4.1** **مبررات اختيار النموذج:** 26](#_Toc220539142)

[**5.4.2** **آلية عمل SegFormer:** 26](#_Toc220539152)

[**5.4.3** **دور SegFormer في المشروع:** 26](#_Toc220539163)

[5.5 نموذج Segment Anything Model (SAM): 27](#_Toc220539164)

[**5.5.1** **دافع استخدام النموذج:** 27](#_Toc220539174)

[**5.5.2** **آلية توظيف SAM ضمن النظام:** 27](#_Toc220539185)

[**5.5.3** **مخرجات النموذج ودورها:** 27](#_Toc220539197)

[5.6 تكامل النماذج وآلية إنتاج النتيجة النهائية: 27](#_Toc220539198)

[**5.6.1** **آلية التفاعل مع المستخدم وتطبيق التعديل البصري:** 28](#_Toc220539209)

[**5.6.2** **التعديل البصري الموجّه (Region-based Editing) :** 29](#_Toc220539221)

[5.7 آلية التقييم والمعايير المعتمدة: 29](#_Toc220539231)

[**2.3.1** **مقياس F1-score1 Macro) وWeighted):** 30](#_Toc220539236)

[**2.3.2** **مصفوفة الالتباس (Confusion Matrix):** 30](#_Toc220539242)

[**2.3.3** **تقرير التصنيف (Classification Report):** 31](#_Toc220539249)

[**2.3.4** **مقياس Mean Intersection over Union (mIoU):** 31](#_Toc220539257)

[**2.3.5** **مقياس** **Mean Accuracy:** 31](#_Toc220539266)

[5.7 مبررات اختيار المنهجية: 32](#_Toc220539267)

[5.8 خلاصة الفصل: 32](#_Toc220539268)

[**الفصل السادس: النتائج والتقييم** 32](#_Toc220539269)

[6.1 مقدّمة الفصل: 32](#_Toc220539271)

[6.2 بيئة التنفيذ ومنهجية التقييم: 33](#_Toc220539272)

[6.3 نتائج نموذج تصنيف نوع الغرفة (Room Classification Model): 33](#_Toc220539273)

[**6.3.1** **إعدادات النموذج وإجراءات التدريب:** 34](#_Toc220539282)

[**6.3.2** **تحليل منحنيات التدريب والتحقق:** 34](#_Toc220539292)

[**6.3.3** **نتائج التقييم النهائي لنموذج تصنيف نوع الغرفة** 35](#_Toc220539303)

[***(Final Evaluation Results – Classification Report)*** 35](#_Toc220539304)

[**6.3.4** **مصفوفة الالتباس (Confusion Matrix):** 37](#_Toc220539316)

[**6.3.5** **مناقشة وخلاصة نموذج تصنيف نوع الغرفة:** 37](#_Toc220539329)

[6.4 نتائج التقسيم الدلالي باستخدام: SegFormer 38](#_Toc220539330)

[6.5 تحسين الأقنعة باستخدام Segment Anything Model (SAM): 39](#_Toc220539331)

[6.6 نتائج التعديل البصري للعناصر المعمارية: 40](#_Toc220539332)

[6.7 نتائج الاختبار على صور واقعية غير موسومة: 41](#_Toc220539333)

[**الفصل السابع: الاستنتاجات والعمل المستقبلي** 42](#_Toc220539334)

[7.1 خلاصة عامة: 42](#_Toc220539336)

[7.2 تحقيق أهداف البحث: 42](#_Toc220539337)

[7.3 حدود النظام الحالي: 42](#_Toc220539338)

[7.4 الآفاق المستقبلية وتطوير العمل: 43](#_Toc220539339)

[7.5 الخلاصة النهائية: 43](#_Toc220539340)

[**الفصل الثامن :المراجع References** 43](#_Toc220539341)

[المراجع الخاصة بالمكتبات والأدوات البرمجية 45](#_Toc220539342)

**المصطلحات والتعاريف:**

|  |  |  |  |
| --- | --- | --- | --- |
| المصطلح التقني | الترجمة العربية | الاختصار | المعنى |
| Artificial Intelligence | الذكاء الاصطناعي | AI | مجال من علوم الحاسوب يهدف إلى تطوير أنظمة قادرة على محاكاة السلوك الذكي البشري مثل التعلّم، الاستدلال، واتخاذ القرار. |
| Machine Learning | تعلّم الآلة | ML | فرع من الذكاء الاصطناعي يركّز على تمكين الأنظمة من التعلّم من البيانات دون برمجة صريحة. |
| Deep Learning | التعلّم العميق | DL | فرع متقدم من تعلّم الآلة يعتمد على شبكات عصبية متعددة الطبقات لاستخلاص تمثيلات معقّدة من البيانات. |
| Neural Network | الشبكة العصبية | NN | نموذج حسابي مستوحى من عمل الدماغ البشري، يتكوّن من عقد مترابطة لمعالجة البيانات. |
| Convolutional Neural Networks | الشبكات العصبية الالتفافية | CNN | نوع من الشبكات العصبية مخصّص لمعالجة الصور، يعتمد على عمليات الالتفاف لاستخلاص السمات البصرية. |
| Feature Extraction | استخراج السمات | — | عملية استخلاص الخصائص المميِّزة من البيانات الخام لاستخدامها في التدريب أو التحليل. |
| Computer Vision | الرؤية الحاسوبية | CV | مجال يهتم بتمكين الحاسوب من فهم وتحليل الصور والمشاهد البصرية. |
| Indoor Scene Analysis | تحليل المشاهد الداخلية | — | عملية فهم مكوّنات الغرف الداخلية والعلاقات المكانية بينها اعتمادًا على الصور. |
| Scene Classification | تصنيف المشهد | — | مهمة تهدف إلى تحديد نوع المشهد العام للصورة مثل غرفة نوم أو مطبخ. |
| Room Classification | تصنيف نوع الغرفة | — | عملية تحديد نوع الغرفة اعتمادًا على السمات البصرية للصورة. |
| Context Awareness | الوعي بالسياق | — | قدرة النموذج على فهم الصورة ضمن إطارها العام وليس فقط العناصر الفردية. |
| Transformer | المحوِّلات | — | بنية شبكات عصبية تعتمد على آلية الانتباه الذاتي لفهم العلاقات السياقية. |
| Self-Attention | الانتباه الذاتي | — | آلية تمكّن النموذج من تقييم أهمية كل جزء من الصورة بالنسبة لبقية الأجزاء. |
| Vision Transformer | محوّل الرؤية | ViT | نموذج Transformer مخصّص لمعالجة الصور عبر تقسيمها إلى رقع. |
| Semantic Segmentation | التقسيم الدلالي | — | مهمة تهدف إلى تصنيف كل بكسل في الصورة ضمن فئة دلالية محددة. |
| Pixel-level Classification | تصنيف على مستوى البكسل | — | إسناد فئة دلالية لكل بكسل في الصورة. |
| Mask | القناع | — | صورة ثنائية تمثل منطقة عنصر معين داخل الصورة. |
| Binary Mask | قناع ثنائي | — | قناع يتكوّن من قيمتين فقط (0 و1) لتمثيل وجود أو غياب العنصر. |
| SegFormer | سيغفورمر | — | نموذج تقسيم دلالي يعتمد على Transformers ويتميّز بالكفاءة والدقة. |
| Segment Anything Model | نموذج تقسيم أي شيء | SAM | نموذج عام لتحسين الأقنعة وتحديد الحدود الهندسية بدقة عالية. |
| Boundary Refinement | تحسين الحدود | — | عملية تهدف إلى تصحيح الحواف الهندسية للأقنعة الناتجة. |
| Geometric Accuracy | الدقة الهندسية | — | مدى تطابق حدود القناع مع الحدود الحقيقية للعنصر في الصورة. |
| Labeled Data | بيانات موسومة | — | بيانات مرفقة بتسميات صحيحة تُستخدم في تدريب النماذج. |
| Unlabeled Data | بيانات غير موسومة | — | بيانات دون تسميات تُستخدم للاختبار أو التعميم. |
| Dataset | مجموعة بيانات | — | مجموعة منظّمة من البيانات تُستخدم في التدريب أو التقييم. |
| NYU Depth V2 | مجموعة بيانات NYU | — | مجموعة بيانات داخلية موسومة تحتوي صور RGB وتسميات دلالية. |
| Places365 | مجموعة بيانات Places365 | — | مجموعة ضخمة من الصور لمشاهد داخلية وخارجية. |
| Ground Truth | الحقيقة الأرضية | GT | القيم الصحيحة المرجعية المستخدمة لتقييم أداء النموذج. |
| Training | التدريب | — | عملية تعلّم النموذج من البيانات الموسومة. |
| Validation | التحقق | — | مرحلة تقييم أداء النموذج أثناء التدريب. |
| Testing | الاختبار | — | مرحلة تقييم النموذج على بيانات غير مستخدمة في التدريب. |
| Accuracy | الدقة | — | نسبة التنبؤات الصحيحة إلى إجمالي التنبؤات. |
| Precision | الدقة الإيجابية | — | نسبة التنبؤات الصحيحة من جميع التنبؤات الإيجابية. |
| Recall | الاسترجاع | — | نسبة العينات الصحيحة المكتشفة من جميع العينات الحقيقية. |
| F1-score | مقياس F1 | — | متوسط توافقي بين Precision وRecall. |
| Confusion Matrix | مصفوفة الالتباس | — | جدول يوضّح التنبؤات الصحيحة والخاطئة لكل فئة. |
| Classification Report | تقرير التصنيف | — | تقرير يعرض مقاييس الأداء لكل فئة على حدة. |
| Mean Intersection over Union | متوسط التقاطع على الاتحاد | mIoU | مقياس أساسي لتقييم جودة التقسيم الدلالي. |
| Mean Accuracy | متوسط الدقة | — | متوسط دقة تصنيف البكسلات عبر جميع الفئات. |
| Visual Editing | التعديل البصري | — | تطبيق تغييرات موضعية على الصورة مثل تغيير لون أو مادة. |
| Surface Restyling | إعادة تصميم السطوح | — | تغيير مظهر الجدران أو الأرضيات مع الحفاظ على الواقعية. |
| User Interface | واجهة المستخدم | UI | الواجهة التي تمكّن المستخدم من التفاعل مع النظام. |
| Realism | الواقعية | — | مدى انسجام التعديل البصري مع خصائص الصورة الأصلية. |

# **الفصل الأول: المقدمة**

## خلفية البحث:

شهد مجال الرؤية الحاسوبية (Computer Vision) تطورًا متسارعًا وملحوظًا خلال السنوات الأخيرة، وذلك نتيجة التقدم الكبير في تقنيات التعلّم العميق (Deep Learning) وما رافقه من تطور في بنى الشبكات العصبية العميقة، ولا سيما النماذج المعتمدة على الشبكات العصبية الالتفافية (CNNs) ثم النماذج الأحدث القائمة على بنية المحوّلات Transformers) )وقد مكّن هذا التقدم الأنظمة الحاسوبية من الانتقال من مجرد معالجة الصور على مستوى السمات البسيطة، إلى فهم أعمق لمحتوى الصورة وتحليلها على مستويات متعددة، تبدأ من استيعاب المشهد العام، ولا تنتهي عند تحديد مكوّناته الدقيقة على مستوى البكسل.

في هذا السياق، يُعد تحليل المشاهد الداخلية، مثل صور الغرف والمساحات المغلقة، من أكثر مهام الرؤية الحاسوبية تعقيدًا مقارنة بالمشاهد الخارجية. ويعود ذلك إلى طبيعة هذه المشاهد التي تتسم بتداخل العناصر المعمارية مع الأثاث، وتشابه الخواص البصرية بين عناصر مختلفة كالجدران والأسقف والأرضيات، إضافة إلى التأثير الكبير لظروف الإضاءة، والظلال، وزوايا التصوير، مما يجعل عملية التمييز الدقيق بين مكونات المشهد تحديًا حقيقيًا على المستويين الدلالي والهندسي.

ومع تزايد الاهتمام بتطبيقات الذكاء الاصطناعي في مجالات التصميم الداخلي، والتسويق العقاري، والواقع المعزز، برزت الحاجة إلى أنظمة قادرة على تحليل صور الغرف وتحويلها إلى تمثيل قابل للتعديل البصري قبل التنفيذ الفعلي على أرض الواقع. إذ تعتمد هذه التطبيقات على إمكانية تغيير عناصر محددة داخل الصورة، مثل لون الجدران أو مادة الأرضيات، بطريقة تحافظ على واقعية المشهد دون التأثير غير المرغوب فيه على العناصر الأخرى.

إن تحقيق هذا النوع من التعديل البصري الموجّه يتطلب وجود نظام قادر على فهم محتوى الصورة فهمًا دقيقًا، بحيث يمكن تحديد موقع وحدود العناصر المستهدفة بدقة عالية. وللوصول إلى هذا المستوى من الفهم، يتم عادةً الاعتماد على مسار معالجة يتألف من مرحلتين أساسيتين ومتكاملتين. تتمثل المرحلة الأولى في تصنيف المشهد الداخلي بهدف تحديد نوع الغرفة، الأمر الذي يساهم في توفير سياق عام للصورة ويساعد في توقع توزيع العناصر المعمارية داخلها. أما المرحلة الثانية فتتمثل في التقسيم الدلالي، والذي يهدف إلى تصنيف كل بكسل في الصورة ضمن فئة دلالية محددة، مما يسمح بفصل العناصر الأساسية داخل المشهد وإنتاج أقنعة تمثل كل عنصر على حدة.

وقد شهد مجال التقسيم الدلالي تطورًا ملحوظًا مع ظهور نماذج حديثة تعتمد على المحوّلات، من أبرزها نموذج SegFormer، الذي أظهر قدرة عالية على تحقيق توازن فعّال بين الكفاءة الحسابية وجودة النتائج، خاصة عند التعامل مع مشاهد معقدة وغنية بالتفاصيل مثل المشاهد الداخلية. ورغم ذلك، فإن الأقنعة الناتجة عن التقسيم الدلالي الأولي قد تعاني من عدم دقة في الحدود الهندسية، أو من وجود تشويش ناتج عن تداخل العناصر أو ضعف التباين البصري.

وللتغلب على هذه الإشكالية، برزت نماذج متقدمة تهدف إلى تحسين الأقنعة الناتجة وزيادة دقتها الهندسية، من بينها نموذج Segment Anything Model (SAM)، الذي يتميز بقدرته على إنتاج أقنعة عالية الدقة وقابلة للتعميم على أنواع متعددة من الصور دون الحاجة إلى تدريب مخصص لكل حالة. ويتيح دمج نتائج التقسيم الدلالي مع نماذج تحسين الأقنعة الوصول إلى تمثيل أدق للعناصر المعمارية داخل المشهد، مما يشكّل خطوة أساسية لدعم تطبيقات التعديل البصري الواقعية في صور الغرف الداخلية

## مشكلة البحث:

على الرغم من التقدم الكبير الذي حققته نماذج تحليل الصور في مجال الرؤية الحاسوبية، لا تزال عملية عزل عناصر محددة داخل صور المشاهد الداخلية تمثل تحديًا بحثيًا قائمًا. إذ تتسم المشاهد الداخلية بتداخل العناصر المعمارية مع الأثاث، وتشابه الخصائص البصرية بين مكوّنات مختلفة مثل الجدران والأرضيات والأسقف، إضافة إلى التأثير الكبير لعوامل الإضاءة والظلال، مما يجعل الفصل الدقيق بين هذه العناصر عملية معقدة على المستويين الدلالي والهندسي.

إن أي عدم دقة في تحديد الحدود الهندسية للعناصر المستهدفة ينعكس سلبًا على موثوقية الأقنعة الناتجة، ويؤدي إلى نتائج غير دقيقة عند استخدامها في تطبيقات لاحقة، الأمر الذي يحدّ من قابلية توظيف هذه الأقنعة في سيناريوهات التعديل البصري الواقعية.

تتمثل المشكلة الأساسية لهذا البحث في كيفية استخراج أقنعة دقيقة وموثوقة للعناصر المعمارية الأساسية داخل الغرفة اعتمادًا على صورة واحدة فقط، بحيث تكون هذه الأقنعة قابلة للاستخدام العملي في مراحل لاحقة مثل التعديل البصري الموجّه. وتزداد هذه الإشكالية تعقيدًا عند عدم الأخذ بعين الاعتبار نوع الغرفة، إذ يختلف توزيع العناصر وخصائصها البصرية بين غرف النوم والمطابخ وصالات المعيشة، مما يجعل الاعتماد على معايير تقسيم ثابتة غير كافٍ لتحقيق نتائج دقيقة وقابلة للتعميم.

وبناءً على ما سبق، تبرز الحاجة إلى اعتماد مسار عمل متكامل يجمع بين فهم المشهد العام للصورة، وتنفيذ التقسيم الدلالي الأولي، ثم تحسين الأقنعة الناتجة للوصول إلى مستوى من الدقة الهندسية والدلالية يتناسب مع متطلبات التطبيقات العملية.

## مسائل البحث:

انطلاقًا من مشكلة البحث المتعلقة بتحليل صور الغرف الداخلية واستخراج أقنعة دقيقة للعناصر المعمارية بهدف دعم التعديل البصري الموجّه، يسعى هذا البحث إلى دراسة ومعالجة المسائل البحثية التالية:

دراسة مدى إمكانية تصنيف نوع الغرفة اعتمادًا على صورة واحدة فقط، وأثر هذا التصنيف في فهم السياق العام للمشهد الداخلي.

دراسة كفاءة نماذج التقسيم الدلالي في استخراج الأقنعة الأولية للعناصر الأساسية داخل الغرفة، مثل الجدران والأرضيات والأسقف.

تحليل أثر تحسين الأقنعة الناتجة باستخدام نماذج متخصصة في تحسين الحدود وتقليل التشويش، ومدى انعكاس ذلك على دقتها الهندسية.

تقييم إمكانية الاستفادة من الأقنعة المحسّنة في تنفيذ تعديلات بصرية موضعية تجريبية (كتغيير اللون أو المادة)، دون اعتبار هذه المرحلة مخرجًا نهائيًا للنظام.

## أثر وأهمية البحث:

تكمن أهمية هذا البحث في تقديمه إطارًا منهجيًا متكاملًا يجمع بين عدة تقنيات حديثة في مجال الرؤية الحاسوبية لمعالجة مشكلة تحليل المشاهد الداخلية، وذلك من خلال الدمج بين فهم المشهد العام، والتقسيم الدلالي، وتحسين الأقنعة الناتجة. ويساهم هذا الدمج في تحسين دقة تمثيل العناصر المعمارية الأساسية داخل صور الغرف الداخلية مقارنة بالاعتماد على مرحلة واحدة فقط، الأمر الذي يعزز موثوقية النتائج الناتجة من الناحية الدلالية والهندسية.

على الصعيد العلمي، يقدّم هذا البحث مساهمة في دراسة أثر تكامل المراحل المختلفة لتحليل الصور على جودة الأقنعة الناتجة، ويبرز أهمية تحسين الحدود الهندسية كعامل أساسي في ربط نتائج التقسيم الدلالي بالتطبيقات العملية. كما يفتح المجال أمام أبحاث مستقبلية تهدف إلى تطوير مسارات معالجة أكثر دقة وقابلية للتعميم عند التعامل مع مشاهد داخلية متنوعة ومعقدة.

أما على الصعيد التطبيقي، فيوفّر هذا العمل أساسًا يمكن البناء عليه لتطوير تطبيقات ذكية في مجالات التصميم الداخلي، والتفاعل البصري، والواقع المعزز، حيث يتيح للمستخدمين إمكانية تصور التغييرات المحتملة على عناصر الغرفة قبل تنفيذها فعليًا. وتبرز أهمية هذا الجانب في تقليل الوقت والجهد اللازمين لاتخاذ قرارات تصميمية، مع الحفاظ على واقعية النتائج البصرية.

إضافة إلى ذلك، يؤكد البحث على الدور المحوري لجودة الأقنعة في نجاح أي نظام يعتمد على التعديل البصري الموجّه، إذ تمثل الأقنعة الدقيقة شرطًا أساسيًا لتفادي الأخطاء البصرية وضمان سلامة التعديلات المطبقة. وبناءً على ذلك، يمكن اعتبار النتائج المتحصلة من هذا البحث نقطة انطلاق لأعمال مستقبلية تشمل دمج واجهات تفاعلية أكثر تطورًا، أو توسيع النظام لدعم تقنيات الواقع المعزز، دون أن يقتصر نطاق الاستفادة على التطبيق الحالي فقط.

# **الفصل الثاني: الدراسة المرجعية للتطبيقات والأبحاث ذات الصلة بالمشروع:**

## الدراسة المرجعية للتطبيقات والبرامج المشابهة:

شهدت السنوات الأخيرة ظهور عدد كبير من التطبيقات المعتمدة على الذكاء الاصطناعي في مجال التصميم الداخلي، والتي تهدف إلى مساعدة المستخدمين على تصور التغييرات المحتملة داخل المساحات الداخلية انطلاقًا من صورة واحدة فقط. وتُعد هذه التطبيقات مرجعًا عمليًا مهمًا للمشروع الحالي، كونها تمثل واقع الاستخدام الفعلي للتقنيات الحديثة، وتُبرز في الوقت نفسه القيود والتحديات المرتبطة بتطبيق التعديل البصري على المشاهد الداخلية.

تناول التطبيق [1] **Home AI – AI Interior Design** فكرة إعادة تصميم الغرفة اعتمادًا على صورة واحدة، حيث يتيح للمستخدم اختيار نوع الغرفة أو الأسلوب التصميمي للحصول على تصورات بصرية بديلة. يركّز هذا التطبيق على سرعة الحصول على النتائج وتعدد الأنماط المقترحة، إلا أنه يعتمد في الغالب على إعادة توليد المشهد بالكامل، مما يحدّ من إمكانية التحكم الدقيق في عناصر محددة مثل الجدران أو الأرضيات.

في السياق نفسه، قدّم التطبيق [2] **RoomGPT** آلية مشابهة تعتمد على توليد تصاميم جديدة للغرف انطلاقًا من الصورة الأصلية، مع التركيز على تنويع الأساليب التصميمية. ورغم فعاليته في إظهار بدائل شكلية متعددة، إلا أن التعديل يتم على مستوى المشهد ككل، دون عزل واضح للأسطح الداخلية، مما قد يؤدي إلى تغييرات غير مرغوبة في عناصر غير مستهدفة.

أما التطبيق [3] **Roomify**، فيعتمد على رفع صورة الغرفة ثم اختيار نوعها والأسلوب المطلوب للحصول على تصميمات مولّدة. ويُظهر هذا التطبيق أهمية مرحلتي تصنيف نوع الغرفة واختيار الأسلوب، إلا أنه لا يوفّر آلية صريحة للتحكم الهندسي الدقيق بحدود الجدران أو الأرضيات قبل تطبيق التعديل.

قدّم التطبيق [4] **DECOZY** مفهوم إعادة التصميم باستخدام الذكاء الاصطناعي مع دعم أنواع متعددة من المساحات الداخلية. ويركّز التطبيق على النتيجة النهائية من حيث الشكل العام، بينما يبقى فصل العناصر الداخلية مرحلة ضمنية غير معروضة للمستخدم، وهو ما يحدّ من إمكانية تقييم دقة التعديل الموضعي.

في التطبيق [5] **DecAI – AI Home Design**، يتم الاعتماد على تحليل الصورة لاختيار نوع الغرفة والأسلوب، ثم توليد اقتراحات تصميمية. وتُظهر هذه المقاربة اعتمادًا واضحًا على فهم السياق العام للمشهد، إلا أن التعديل يبقى موجّهًا للمشهد ككل أكثر من كونه تعديلًا دقيقًا لسطوح محددة.

أما التطبيق [6] **Fotor – AI Interior Design**، فيندرج ضمن فئة الأدوات العامة التي تتيح تعديل الصور باستخدام تقنيات الذكاء الاصطناعي، بما في ذلك تعديل أجزاء من الصورة. ورغم قدرته على إنتاج نتائج جذابة بصريًا، إلا أن دقة التعديل تعتمد بشكل كبير على جودة الفصل بين العناصر، والتي لا تكون دائمًا مخصصة لأسطح الغرف الداخلية.

يوفّر التطبيق [7] **mnml.ai – Interior AI Render** أدوات لتوليد تصورات متعددة للمساحات الداخلية، ويُستخدم غالبًا من قبل المصممين لتسريع عملية إنتاج البدائل التصميمية. غير أن هذا النوع من التطبيقات يركّز على توليد التنويعات البصرية أكثر من التركيز على مسار واضح لعزل الأسطح وتحسين أقنعتها قبل التعديل.

ومن التطبيقات التي تقترب بشكل مباشر من فكرة المشروع، يأتي التطبيق [8] **REimagineHome.ai**، الذي يقدّم ميزات مخصّصة لإعادة تلوين الجدران أو تغيير خامات الأرضيات. ويُظهر هذا التطبيق بوضوح الحاجة العملية إلى التعديل الموضعي للأسطح، كما يبرز في الوقت نفسه أهمية دقة الأقنعة المستخدمة لعزل هذه الأسطح عن بقية عناصر المشهد.

في مجال التفاعل والواقع المعزز، يقدّم التطبيق [9] **DecorMatters** تجربة تفاعلية لتجربة عناصر التصميم داخل الغرف باستخدام تقنيات AR، مع التركيز على وضع الأثاث والديكور. ورغم اختلاف هدفه الأساسي عن مشروعنا، إلا أنه يبرز أهمية الإقناع البصري والتفاعل في تطبيقات التصميم الداخلي.

كما يقدّم النظام [10] **IKEA Kreativ** تجربة متقدمة تعتمد على مسح المشهد وبناء تمثيل أقرب إلى ثلاثي الأبعاد، مما يتيح تعديل المساحة بشكل أكثر واقعية. وتُذكر هذه التجربة كاتجاه متقدم في المجال، حيث تصبح دقة الفصل الهندسي وفهم البنية المكانية عاملين أساسيين لنجاح التعديل البصري.

تشير هذه التطبيقات مجتمعةً إلى أن السوق يتجه نحو حلول تتيح للمستخدم تصور التعديلات قبل التنفيذ، إلا أن معظمها لا يقدّم مسارًا واضحًا وقابلًا للتقييم لعزل الأسطح الداخلية وتحسين أقنعتها قبل التعديل. ومن هنا، يتموضع مشروعنا كحل يهدف إلى سد هذه الفجوة من خلال اعتماد مسار منهجي يبدأ بفهم المشهد الداخلي، ثم استخراج الأقنعة الأولية، وتحسينها، قبل تطبيق أي تعديل بصري موجّه.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| المرجع | السنة | |  | | --- | | **اسم التطبيق** | | |  | | --- | | **نوع الإدخال** | | آلية العمل الأساسية | |  | | --- | | **نوع التعديل** | | ملاحظات مرتبطة بالمشروع |
| [1] | 2022 | Home AI -AI | صورة غرفة واحدة | تحليل المشهد + اختيار ستايل | إعادة تصميم كاملة | لا يوفّر تحكمًا دقيقًا بعزل الجدران أو الأرضيات |
| [2] | 2023 | RoomGPT | صورة واحدة | توليد تصاميم حسب الستايل | إعادة توليد المشهد | التركيز على النتيجة النهائية أكثر من دقة الأقنعة |
| [3] | 2023 | Roomify | صورة داخلية | تصنيف نوع الغرفة + توليد بدائل | إعادة تصميم عام | غياب آلية صريحة للتحكم الهندسي بالأسطح |
| [4] | 2023 | DECOZY | صورة غرفة | توليد تصميم بالاعتماد على الذكاء الاصطناعي | إعادة تصميم شامل | فصل العناصر يتم ضمنيًا وغير قابل للتقييم |
| [5] | 2023 | DecAI – AI Home Design | صورة | تحليل المشهد + اختيار نوع المساحة | تعديل بصري عام | يعتمد على السياق العام دون عزل دقيق للأسطح |
| [6] | 2023 | Fotor – AI Interior Design | صورة أو مخطط | أدوات تعديل وتوليد عامة | تعديل جزئي أو كامل | غير مخصص لأسطح الغرف الداخلية |
| [7] | 2023 | mnml.ai – Interior | صورة غرفة | توليد تنويعات تصميمية | إعادة تصور المساحة | موجّه لتسريع العمل وليس للتحرير الموضعي |
| [8] | 2024 | REimagineHome.ai | صورة + AR | Surface Restyling | تعديل الجدران/الأرضيات | الأقرب لفكرة التعديل الموضعي للأسطح |
| [9] | 2024 | DecorMatters | صورة + مسح مشهد | تجربة تفاعلية بالواقع المعزز | وضع عناصر وأثاث | يركز على التفاعل أكثر من تعديل الأسطح |
| [10] | 2024 | IKEA Kreativ | صورة + مسح مشهد | Scene Scanning + تمثيل شبه ثلاثي الأبعاد | تعديل شامل للمساحة | يبرز أهمية الدقة الهندسية كاتجاه مستقبلي |

## الدراسة المرجعية للأبحاث والأوراق العلمية:

تناولت العديد من الدراسات البحثية مسألة فهم المشاهد الداخلية باستخدام تقنيات التعلّم العميق، حيث انطلقت الأعمال الأولى من التركيز على تصنيف نوع الغرفة اعتمادًا على السمات العامة للصورة، بهدف استنتاج السياق العام للمشهد الداخلي. وقد اعتمدت هذه الدراسات بشكل أساسي على مجموعات بيانات تحتوي على عدد كبير من الصور الداخلية الملتقطة من بيئات واقعية متنوعة، تختلف من حيث الإضاءة، زوايا التصوير، توزيع الأثاث، وأنماط التصميم. وتضم هذه البيانات فئات متعددة مثل غرف النوم، المطابخ، غرف المعيشة، المكاتب المنزلية والمداخل، مما يجعل مهمة التصنيف أكثر تعقيدًا بسبب التشابه البصري بين بعض الفئات.

في هذا السياق، ركّزت الدراسة [1] على تحسين دقة تصنيف المشاهد الداخلية من خلال اعتماد بنية Dual-Stream Deep Learning، حيث يعالج المسار الأول الصورة على مستوى السياق العام، في حين يركّز المسار الثاني على التفاصيل المحلية الدقيقة. تم تدريب النموذج على مجموعة بيانات داخلية كبيرة تضم آلاف الصور، وتم تقييم الأداء باستخدام مقياس الدقة (Accuracy) بوصفه المقياس الأكثر شيوعًا في مهام تصنيف المشاهد. وأظهرت النتائج تحسّنًا ملحوظًا مقارنة بالنماذج ذات المسار الواحد، حيث حقق النموذج دقة تصنيف تجاوزت 85%، مما يدل على أن دمج المعلومات العامة والتفصيلية يساهم بشكل مباشر في تحسين فهم المشاهد الداخلية.

وفي دراسة لاحقة [2]، تم الانتقال من الاعتماد على الصورة الكاملة فقط إلى استغلال المعلومات الدلالية المستخرجة من التقسيم الدلالي. استخدمت الدراسة عدة مجموعات بيانات داخلية متنوعة، حيث تم استخراج ميزات إضافية من أقنعة تمثل مواقع الأجسام والعناصر داخل الغرفة، مثل الأثاث والجدران والأرضيات، ثم دمج هذه الميزات مع الميزات العامة للصورة ضمن نموذج تصنيف واحد. أظهرت النتائج، التي تم تقييمها باستخدام Accuracy، تحسّنًا في التمييز بين الغرف المتشابهة بصريًا، مع وصول الدقة إلى ما يقارب 90%، مما يؤكد أهمية المعلومات الدلالية في دعم مهام تصنيف المشاهد الداخلية.

مع تطور الاهتمام بتحليل مكوّنات المشهد الداخلي، ركّزت دراسات أخرى [3] على التقسيم الدلالي للمشاهد الداخلية باستخدام نماذج قائمة على المحوّلات، مثل SegFormer. اعتمدت هذه الدراسات على مجموعات بيانات قياسية واسعة الانتشار، أبرزها ADE20K، والتي تحتوي على عشرات الآلاف من الصور الداخلية والخارجية الموسومة على مستوى البكسل، وتشمل عددًا كبيرًا من الفئات الدلالية. تم تقييم الأداء باستخدام mIoU، وأظهرت النتائج تحقيق قيم mIoU تراوحت تقريبًا بين 37% و55% تبعًا لنسخة النموذج وإعدادات التدريب، مع كفاءة حسابية محسّنة مقارنة بنماذج التقسيم التقليدية، مما يجعل هذا النوع من النماذج مناسبًا للتطبيقات العملية واسعة النطاق.

إلا أن التقسيم الدلالي التقليدي قد يعاني من ضعف في دقة الحدود الهندسية بين العناصر المتجاورة داخل المشهد، مثل التداخل بين الجدار والأرضية أو بين الأثاث والأسطح. لمعالجة هذه المشكلة، تناولت دراسات أحدث [4] استخدام نماذج تحسين الأقنعة مثل Segment Anything Model (SAM). اعتمدت هذه الدراسات على مجموعات بيانات ضخمة جدًا ومتنوعة تضم صورًا من مجالات متعددة، دون الاقتصار على المشاهد الداخلية فقط. تم تقييم الأداء باستخدام IoU إلى جانب التقييم البصري، وأظهرت النتائج تحسّنًا ملحوظًا في قيم IoU وجودة الحدود مقارنة بالأقنعة الناتجة مباشرة من نماذج التقسيم الدلالي، دون تحديد قيمة رقمية واحدة ثابتة بسبب اختلاف سيناريوهات التقييم.

كما ناقشت دراسات أخرى [5] إمكانية دمج معلومات العمق أو العمق الزائف (Pseudo Depth) لتحسين فهم البنية المكانية للمشاهد الداخلية. استخدمت هذه الدراسات بيانات تحتوي على صور RGB مرفقة بخرائط عمق حقيقية أو مولّدة، وتم تقييم الأداء باستخدام mIoU. وأظهرت النتائج أن إدخال معلومات العمق أدى إلى تحسّن في قيم mIoU بنسبة تتراوح تقريبًا بين 3% و6% مقارنة بالنماذج التي تعتمد على الصور ثنائية الأبعاد فقط، خاصة في فصل الأسطح الرئيسية مثل الأرضيات والجدران.

إضافة إلى ذلك، تناولت أبحاث حديثة [6] و[7] مسألة اختيار الألوان وتوليد لوحات الألوان في سياق التصميم الداخلي. اعتمدت هذه الدراسات على مجموعات بيانات تصميمية وصور داخلية رقمية، وركّزت على الجوانب الجمالية والمفاهيمية بدلًا من المقاييس الرقمية الصارمة. لذلك تم تقييم النتائج غالبًا باستخدام التقييم النوعي (Qualitative Evaluation)، الذي يعتمد على مدى انسجام الألوان مع السياق العام للمشهد ورضا المستخدم، وأكدت هذه الدراسات أن اللون عنصر أساسي في التصميم الداخلي ويرتبط ارتباطًا وثيقًا بوظيفة الغرفة وطبيعتها.

تشير هذه الدراسات مجتمعةً إلى أن معالجة المشاهد الداخلية بشكل فعّال تتطلب نهجًا متعدد المراحل، يبدأ بفهم السياق العام للمشهد عبر تصنيف نوع الغرفة، ثم تحليل المكوّنات الأساسية باستخدام التقسيم الدلالي، يلي ذلك تحسين الأقنعة ودقة الحدود، وصولًا إلى دعم قرارات التصميم مثل اختيار المواد والألوان. ويندرج مشروعنا ضمن هذا التوجه من خلال اعتماد مسار متكامل يجمع بين تصنيف المشهد، والتقسيم الدلالي، وتحسين الأقنعة، بما يخدم هدف التعديل البصري الموجّه للأسطح الداخلية.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| المرجع | السنة | |  | | --- | | **عنوان الدراسة** | | |  | | --- | | **نوع االبيانات** | | حجم البيانات | |  | | --- | | **المهمة** | | النموذج المستخدم | المقاييس | النتائج |
| [1] | 2024 | Indoor Scene Classification through Dual-Stream DL | صور داخلية | آلاف الصور | |  | | --- | | تصنيف نوع الغرفة | | Dual-Stream DL | Accuracy | تحسن ملحوظ مقارنة بمسار واحد (Accuracy > 85%) |
| [2] | 2024 | Exploiting Object-  based and Segmentation-based Features | صور داخلية | عدة مجموعات | تصنيف المشاهد | CNN + Segmentation | Accuracy | تحسن التمييز بين الغرف المتشابهة (Accuracy ≈ 90%) |
| [3] | 2021 | SegFormer | صور داخلية/خارجية | ADE20K وغيرها | تقسيم دلالي | Transformer | mIoU | نتائج دقيقة مع كلفة حسابية منخفضة (mIoU ≈ 37–55%) |
| [4] | 2023 | |  | | --- | | Segment Anything | | صور عامة | واسع جدًا | تحسين الأقنعة | SAM | |  | | --- | | IoU | | تحسين دقة الحدود بشكل واضح تحسن ملحوظ في IoU)**)** |
| [5] | 2025 | PDDM: Pseudo Depth Diffusion Model | RGB + عمق زائف | مشاهد داخلية | تقسيم دلالي | Diffusion + CNN | mIoU | تحسين دقة فصل الأسطح (تحسن mIoU بنسبة 3–6%) |
| [6] | 2025 | GenColor | صور تصميم | غير مذكور | ربط اللون بالمفهوم | Generative Model | Qualitative | تحسين اختيار الألوان (تقييم نوعي إيجابي) |
| [7] | 2024 | Color Palette Generation Review | صور رقمية | متعددة | توليد لوحات ألوان | Methods Review | — | تحليل شامل لأساليب اختيار الألوان (Qualitative Evaluation) |

# **الفصل الثالث: الدراسة النظرية**



## مقدّمة:

يُعدّ هذا الفصل الركيزة النظرية الأساسية التي يقوم عليها هذا البحث، إذ يهدف إلى تقديم عرض علمي متكامل للمفاهيم والنماذج الحديثة في مجال الذكاء الاصطناعي والرؤية الحاسوبية، مع التركيز على التقنيات التي تُستخدم في تحليل المشاهد الداخلية وفهم مكوّناتها. وتكمن أهمية هذا الفصل في كونه يوفّر الإطار المفاهيمي الذي يفسّر الأسس العلمية للمنهجية العملية المعتمدة لاحقًا، ويضمن وجود ترابط منطقي بين الجانب النظري والتطبيقي للبحث.

شهد مجال الرؤية الحاسوبية تطورًا ملحوظًا خلال العقود الأخيرة، حيث انتقلت الأنظمة الحاسوبية من الاعتماد على أساليب تقليدية قائمة على القواعد والسمات المصمّمة يدويًا، إلى نماذج ذكية قادرة على التعلّم التلقائي من البيانات. وقد أظهرت الأساليب التقليدية محدودية واضحة عند التعامل مع الصور الواقعية المعقّدة، لا سيما المشاهد الداخلية التي تتسم بتنوّع عناصرها، وتداخل مكوّناتها، واختلاف ظروف الإضاءة والمنظور.

مع ظهور التعلّم العميق، أصبح بالإمكان تطوير نماذج قادرة على استخلاص التمثيلات البصرية تلقائيًا، والتعامل مع التعقيد البصري دون الحاجة إلى تدخل بشري مباشر في تصميم السمات. وقد شكّل هذا التحول نقطة انطلاق لتطوير نماذج أكثر دقة ومرونة، قادرة على فهم المشاهد الداخلية وتحليل بنيتها المكانية والدلالية بشكل أكثر شمولية.

يستعرض هذا الفصل بدايةً التعلّم العميق بوصفه الأساس النظري العام لمعالجة الصور الحديثة، ثم يتناول الشبكات العصبية الالتفافية التي مثّلت حجر الأساس في استخراج السمات البصرية. بعد ذلك، يناقش الفصل التحول النوعي الذي أحدثته نماذج Transformers في إدراك السياق العام للمشهد. ويُختتم الجزء النظري بتوضيح الدور الذي تلعبه هذه النماذج في تمهيد الطريق لمهام أكثر تقدّمًا، مثل التقسيم الدلالي وتحسين الدقة الهندسية للأقنعة.

## أساسيات التعلّم العميق:



### **مفهوم التعلّم العميق ودوافع استخدامه:**

يُعدّ التعلّم العميق (Deep Learning) أحد الفروع المتقدمة لتعلّم الآلة، ويعتمد على شبكات عصبية اصطناعية متعددة الطبقات تهدف إلى محاكاة آلية معالجة المعلومات في الدماغ البشري. تتميّز هذه الشبكات بقدرتها على تعلّم تمثيلات هرمية للبيانات، حيث تنتقل من استخراج السمات البسيطة إلى تمثيل المفاهيم المجرّدة والمعقّدة.جاء ظهور التعلّم العميق نتيجةً للحاجة إلى نماذج قادرة على التعامل مع البيانات غير المهيكلة، مثل الصور والفيديوهات، والتي يصعب توصيفها باستخدام قواعد ثابتة أو سمات مصمّمة يدويًا. ففي مجال الرؤية الحاسوبية، تحتوي الصور على عدد كبير من العلاقات المكانية والبصرية المتشابكة، مثل التدرّجات اللونية، والأنماط الهندسية، والعلاقات بين العناصر داخل المشهد الواحد.أتاح التعلّم العميق للنماذج القدرة على تعلّم هذه العلاقات تلقائيًا من البيانات، دون الحاجة إلى تدخل مباشر من المستخدم في تحديد السمات ذات الصلة. وقد ساهم ذلك في تحقيق قفزة نوعية في أداء الأنظمة الحاسوبية في مهام متعددة، مثل تصنيف الصور، واكتشاف الأجسام، وتحليل المشاهد المعقّدة. وبناءً على ذلك، أصبح التعلّم العميق الأساس النظري الذي تقوم عليه معظم تطبيقات الرؤية الحاسوبية الحديثة.



### **آلية التعلّم داخل الشبكات العميقة:**

تعتمد عملية التعلّم في الشبكات العصبية العميقة على تمرير البيانات عبر سلسلة من الطبقات المتتالية، حيث تقوم كل طبقة بإجراء تحويل رياضي على المدخلات وإنتاج تمثيل جديد أكثر تجريدًا. في سياق معالجة الصور، تتعلّم الطبقات الأولى السمات البصرية البسيطة مثل الحواف والاتجاهات، بينما تتجه الطبقات الأعمق إلى تمثيل الأشكال المعقّدة والعلاقات بين عناصر الصورة.

يُقاس أداء الشبكة باستخدام دالة خسارة تعبّر عن الفرق بين المخرجات المتوقعة والقيم الحقيقية. ويتم تحديث أوزان الشبكة باستخدام خوارزمية الانتشار العكسي، التي تعتمد على حساب مشتقات دالة الخسارة بالنسبة للأوزان وتعديلها تدريجيًا بهدف تقليل الخطأ. وتُكرَّر هذه العملية عبر عدد كبير من دورات التدريب، مما يسمح للنموذج بتحسين أدائه تدريجيًا.

تتيح هذه الآلية للشبكات العميقة تعلّم أنماط معقّدة من البيانات دون الحاجة إلى تدخل بشري مباشر، وهو ما يفسّر النجاح الكبير الذي حققته هذه النماذج في تحليل الصور الواقعية وفهم المشاهد الداخلية.

## الشبكات العصبية الالتفافية (CNNs):



### **بنية الشبكات الالتفافية وأهميتها:**

ظهرت الشبكات العصبية الالتفافية (Convolutional Neural Networks) كحلّ فعّال لمعالجة الصور الرقمية، بعد أن أثبتت الشبكات العصبية التقليدية محدوديتها في التعامل مع البيانات ذات الأبعاد العالية، مثل الصور، والتي تحتوي على عدد كبير من البكسلات والعلاقات المكانية المعقّدة. فقد كانت الشبكات التقليدية تتطلّب عددًا هائلًا من المعاملات عند معالجة الصور مباشرة، مما يؤدي إلى زيادة التعقيد الحسابي وصعوبة التدريب.

تعتمد الشبكات العصبية الالتفافية على مبدأ الاستفادة من البنية المكانية للصورة، وذلك من خلال تطبيق عمليات الالتفاف باستخدام مرشّحات صغيرة الحجم تتحرك عبر الصورة لاستخلاص السمات البصرية المحلية. تسمح هذه المرشّحات باكتشاف الأنماط الأساسية، مثل الحواف، الزوايا، والأنسجة، والتي تُعدّ عناصر أولية لفهم المحتوى البصري للصورة.

تتكوّن البنية النموذجية للشبكات الالتفافية من عدة طبقات متتابعة، تبدأ بطبقات التفاف مسؤولة عن استخلاص السمات، تليها طبقات تجميع تهدف إلى تقليل الأبعاد المكانية للبيانات مع الحفاظ على أهم المعلومات المستخرجة. تسهم هذه الطبقات في تحسين الكفاءة الحسابية وتقليل الحساسية للتغيّرات الطفيفة في موضع العناصر داخل الصورة. وفي المراحل اللاحقة، تُستخدم الطبقات كاملة الاتصال لمعالجة التمثيل النهائي واتخاذ القرار المناسب.

تُعدّ خاصية مشاركة الأوزان من أبرز مميزات الشبكات العصبية الالتفافية، حيث يُستخدم نفس المرشّح عبر مناطق مختلفة من الصورة، مما يؤدي إلى تقليل عدد المعاملات المطلوبة بشكل كبير مقارنةً بالشبكات التقليدية. ونتيجة لذلك، تتميّز الشبكات الالتفافية بقدرتها على معالجة الصور ذات الدقة العالية بكفاءة، الأمر الذي جعلها الأساس في معظم تطبيقات الرؤية الحاسوبية الحديثة.



### **دور الشبكات الالتفافية في فهم المشاهد الداخلية:**

تلعب الشبكات العصبية الالتفافية دورًا محوريًا في فهم وتحليل المشاهد الداخلية، نظرًا لقدرتها على استخلاص السمات البصرية المرتبطة بالعناصر المعمارية الأساسية داخل الصورة. ففي المشاهد الداخلية، تتداخل عناصر متعددة مثل الجدران، الأرضيات، الأسقف، والأثاث، مما يتطلّب نماذج قادرة على تحليل التفاصيل الدقيقة والتمييز بين هذه العناصر بدقة.

تمكّن الشبكات الالتفافية من التعرّف على العناصر المعمارية الرئيسة من خلال تحليل الخصائص البصرية التي تميّز كل عنصر، مثل الاتجاهات الهندسية، القوام السطحي، والتباين اللوني. كما تساهم في التعرّف على مكوّنات الأثاث المختلفة وفهم توزيعها المكاني داخل الغرفة، وهو ما يساعد على بناء تمثيل بصري أوّلي لبنية المشهد الداخلي.

تعتمد هذه الشبكات على تمثيل هرمي للمعلومات، حيث تستخلص الطبقات الأولى السمات البسيطة، بينما تتعلّم الطبقات الأعمق أنماطًا أكثر تعقيدًا تعبّر عن الأشكال والعلاقات بين العناصر. ويسمح هذا التدرّج الهرمي للنموذج بفهم تفاصيل المشهد على مستويات متعددة، بدءًا من التفاصيل الدقيقة وصولًا إلى البنية العامة للصورة.

ورغم هذه القدرات، فإن الشبكات العصبية الالتفافية تعتمد بشكل أساسي على المعالجة المحلية، إذ يكون مجال الرؤية لكل مرشّح محدودًا بنطاق صغير من الصورة. ونتيجة لذلك، قد تواجه هذه الشبكات صعوبة في إدراك العلاقات البعيدة بين عناصر المشهد، مثل التناسق العام بين الأرضية وبقية مكوّنات الغرفة أو العلاقة الشاملة بين توزيع الأثاث والمساحة المتاحة. وقد شكّل هذا القيد أحد أبرز التحديات في تحليل المشاهد الداخلية المعقّدة، ومهّد الطريق لظهور نماذج أكثر تقدّمًا تعتمد على فهم السياق العام للصورة، مثل نماذج Transformers .

## نماذج Transformers في الرؤية الحاسوبية:



### **التحول نحو النماذج السياقية:**

أحدث إدخال نماذج  **Transformers** تحولًا جذريًا في مجال الرؤية الحاسوبية، بعد أن أثبتت هذه البنى نجاحًا كبيرًا في معالجة اللغة الطبيعية. جاء هذا التحول استجابةً للقيود التي تعاني منها النماذج التقليدية المعتمدة على المعالجة المحلية، وعلى رأسها الشبكات العصبية الالتفافية، والتي تركز بشكل أساسي على السمات القريبة ضمن نطاق محدود من الصورة.تعتمد نماذج Transformers في الرؤية الحاسوبية على إعادة تمثيل الصورة بطريقة مختلفة، حيث يتم تقسيمها إلى رقع صغيرة (Patches)، ويُحوَّل كل جزء منها إلى تمثيل عددي يُعالج ضمن تسلسل من البيانات. يسمح هذا الأسلوب للنموذج بالتعامل مع الصورة بوصفها كيانًا متكاملًا، بدلًا من الاقتصار على تحليل المناطق المحلية بشكل منفصل، كما هو الحال في الشبكات الالتفافية.يساهم هذا التمثيل في تمكين النموذج من إدراك العلاقات السياقية بين عناصر المشهد المختلفة، مثل العلاقة بين الأرضية والجدران أو بين توزيع الأثاث والمساحة العامة للغرفة. ويُعدّ هذا الإدراك السياقي عنصرًا بالغ الأهمية في تحليل المشاهد الداخلية، التي تتسم بتداخل العناصر وتنوّعها، حيث لا يمكن فهم كل عنصر بمعزل عن السياق العام الذي ينتمي إليه.من خلال هذا التحول من الفهم المحلي إلى الفهم السياقي، أصبحت نماذج Transformers قادرة على تقديم تمثيل بصري أكثر شمولية للمشهد الداخلي، يجمع بين التفاصيل الدقيقة والعلاقات العامة بين العناصر. وقد مهّد هذا التطور الطريق لاستخدام هذه النماذج في مهام متقدمة، مثل التقسيم الدلالي الدقيق وتحليل البنية المعمارية للمشاهد الداخلية المعقّدة.



### **آلية الانتباه الذاتي ودورها:**

تعتمد نماذج Transformers على آلية الانتباه الذاتي Self-Attention))، والتي تُعدّ المكوّن الأساسي الذي يميّز هذه النماذج عن الأساليب التقليدية في معالجة الصور. تهدف هذه الآلية إلى تمكين النموذج من تحليل العلاقات بين جميع أجزاء الصورة في الوقت نفسه، بدلًا من الاعتماد على نطاق محلي محدود.

تقوم آلية الانتباه الذاتي بحساب درجة الأهمية النسبية لكل جزء من الصورة بالنسبة لبقية الأجزاء، وذلك من خلال تمثيل كل رقعة من الصورة ضمن فضاء عددي، ثم قياس مدى ارتباطها ببقية الرقع. وبناءً على هذه العلاقات، يتم تخصيص أوزان مختلفة لكل جزء أثناء عملية المعالجة، بحيث تحظى المناطق الأكثر تأثيرًا في فهم المشهد بوزن أعلى.

تسمح هذه الآلية للنموذج بفهم العلاقات بين العناصر المتباعدة مكانيًا داخل الصورة، مثل ربط الأرضية بالجدران أو إدراك العلاقة بين مواقع الأثاث المختلفة ضمن الغرفة الواحدة. ويُعدّ هذا الأمر بالغ الأهمية في المشاهد الداخلية، حيث تعتمد دقة الفهم على إدراك التناسق العام بين العناصر وليس فقط على تفاصيلها المحلية.

تعتمد آلية الانتباه الذاتي على عمليات رياضية تُستخدم لإعادة تشكيل التمثيل البصري للصورة بطريقة تُبرز العلاقات الدلالية والسياقية بين عناصرها المختلفة. ونتيجة لذلك، يصبح النموذج قادرًا على تجاوز القيود المرتبطة بالمعالجة المحلية، وتقديم تمثيل بصري أكثر شمولية ووعيًا بالسياق العام للصورة.

يساهم هذا الأسلوب في تحسين أداء النماذج في المهام التي تتطلب فهمًا دقيقًا للعلاقات المكانية المعقّدة، مثل تحليل المشاهد الداخلية والتقسيم الدلالي، حيث يكون إدراك السياق العام وتكامل العناصر شرطًا أساسيًا للحصول على نتائج دقيقة وموثوقة.

## التقسيم الدلالي للصور:



### **مفهوم التقسيم الدلالي:**

يُعدّ التقسيم الدلالي (Semantic Segmentation) من المهام المتقدمة في مجال الرؤية الحاسوبية، ويهدف إلى تصنيف كل بكسل في الصورة إلى فئة دلالية محددة، بحيث يتم إسناد تسمية دلالية لكل نقطة ضمن الصورة. وعلى عكس مهام التصنيف التقليدي أو اكتشاف الأجسام، التي تركز على الصورة ككل أو على كائنات منفصلة، يوفّر التقسيم الدلالي تمثيلًا تفصيليًا للمشهد على مستوى البكسل، مما يسمح بفهم أدق لبنية الصورة ومكوّناتها.ينتج عن عملية التقسيم الدلالي خريطة دلالية توضّح مواقع العناصر المختلفة داخل المشهد وحدودها التقريبية، مثل الجدران، الأرضيات، الأسقف، والأجسام الأخرى. ويُعدّ هذا النوع من التمثيل أساسيًا في التطبيقات التي تتطلب دقة مكانية عالية، إذ لا يكتفي بتحديد وجود العنصر، بل يحدّد امتداده المكاني داخل الصورة.تُستخدم تقنيات التقسيم الدلالي على نطاق واسع في مجالات متعددة، مثل التطبيقات الطبية لتحديد الأنسجة والأعضاء، والقيادة الذاتية لفهم البيئة المحيطة بالمركبة، والتطبيقات المعمارية والهندسية لتحليل المساحات الداخلية والخارجية. ويعكس هذا الانتشار الواسع الأهمية الكبيرة لهذه التقنية في توفير فهم عميق ومفصّل للمشاهد البصرية المعقّدة



### **أهمية التقسيم الدلالي في المشروع:**

تكتسب تقنية التقسيم الدلالي أهمية خاصة في تحليل المشاهد الداخلية، نظرًا للحاجة إلى تحديد العناصر المعمارية بشكل واضح ودقيق. ففي البيئات الداخلية، تتداخل العناصر المختلفة ضمن مساحة محدودة، مما يجعل من الصعب إجراء أي تعديل بصري أو تحليل هندسي دون وجود تمثيل دقيق يحدّد مواقع هذه العناصر وحدودها.

في سياق هذا البحث، يمثّل التقسيم الدلالي الأساس الذي يُبنى عليه التعديل البصري داخل النظام المقترح، إذ يسمح بتحديد المناطق المستهدفة للتعديل على مستوى البكسل، مثل الأرضيات أو الجدران، دون التأثير على بقية عناصر المشهد. ويُعدّ هذا المستوى من الدقة شرطًا أساسيًا لتحقيق نتائج واقعية تحافظ على انسجام الصورة الأصلية.

ومع ذلك، فإن نماذج التقسيم الدلالي التقليدية قد تعاني من بعض القيود، لا سيما فيما يتعلق بدقة الحواف الفاصلة بين العناصر، خاصة في الصور الواقعية التي تتسم بتدرّجات لونية معقّدة أو حدود غير واضحة. وقد تؤدي هذه المشكلة إلى أقنعة غير دقيقة، مما ينعكس سلبًا على جودة التعديلات البصرية اللاحقة. وقد شكّل هذا التحدي دافعًا للبحث عن حلول تكميلية قادرة على تحسين الدقة الهندسية للأقنعة الناتجة.

## **نموذج** :Segment Anything Model (SAM)



### **لمحة عامة عن النموذج:**

يُعدّ نموذج Segment Anything Model (SAM) من النماذج الحديثة التي طُوّرت بهدف توفير إطار عام وقابل للتعميم لعملية التقسيم البصري. يختلف هذا النموذج عن النماذج التقليدية في كونه لا يقتصر على فئات دلالية محددة أو حالات استخدام معيّنة، بل يهدف إلى التعامل مع أنواع متعددة من الصور والعناصر دون الحاجة إلى تدريب مخصّص لكل تطبيق على حدة.

يعتمد نموذج SAM على تدريبه باستخدام مجموعات بيانات ضخمة ومتنوّعة، مما أكسبه قدرة عالية على التعميم والتعامل مع مشاهد بصرية مختلفة. وقد تم تصميمه ليكون نموذجًا عامًا للتقسيم، قادرًا على توليد أقنعة دقيقة لمختلف العناصر داخل الصورة بناءً على مدخلات مرنة، مثل نقاط إرشادية أو أقنعة أولية.

جاء تطوير هذا النموذج استجابةً للحاجة إلى أنظمة تقسيم أكثر مرونة، قادرة على التكيّف مع سيناريوهات جديدة دون إعادة تدريب مكلفة. ويُعدّ هذا التوجّه مهمًا بشكل خاص في التطبيقات العملية، حيث تتغيّر طبيعة الصور والبيئات بشكل مستمر.



### **دوره في تحسين الدقة الهندسية:**

تكمن أهمية نموذج SAM في قدرته العالية على تحديد الحدود الهندسية للعناصر داخل الصورة بدقة كبيرة. ففي حين توفّر نماذج التقسيم الدلالي تمثيلًا جيدًا من الناحية الدلالية، إلا أن دقة الحواف الفاصلة بين العناصر قد تكون محدودة، خاصة في الصور الواقعية ذات التفاصيل المعقّدة.يأتي نموذج SAM ليكمل هذه النتائج من خلال تحسين الحواف وتوليد أقنعة أكثر دقة من الناحية الهندسية، مما يجعلها مناسبة للتطبيقات البصرية العملية التي تتطلب مستوى عاليًا من الواقعية. إذ يساهم النموذج في تقليل التداخل غير المرغوب بين العناصر، والحفاظ على الحدود الطبيعية بينها.وبذلك، يؤدي نموذج SAM دورًا تكميليًا ضمن النظام المقترح، حيث يجمع بين الفهم الدلالي الذي توفّره نماذج التعلّم العميق، والدقة الهندسية اللازمة لتحقيق نتائج بصرية متناسقة وواقعية. ويُعدّ هذا التكامل بين النماذج الدلالية والنماذج الهندسية عنصرًا أساسيًا في تحسين جودة التعديلات البصرية في المشاهد الداخلية.

## تكامل المفاهيم ضمن النظام المقترح:

تُظهر الدراسة النظرية للتقنيات السابقة أن كل مفهوم من المفاهيم المطروحة لا يعمل بشكل مستقل، بل يندمج ضمن إطار متكامل يهدف إلى تحقيق فهم شامل ودقيق للمشاهد الداخلية. إذ يشكّل التعلّم العميق الأساس النظري العام الذي تقوم عليه جميع النماذج المستخدمة، موفّرًا القدرة على تعلّم التمثيلات البصرية المعقّدة من البيانات دون الحاجة إلى تصميم السمات يدويًا.

ضمن هذا الإطار، تؤدي الشبكات العصبية الالتفافية دورًا محوريًا في استخلاص السمات البصرية المحلية المرتبطة بالعناصر المعمارية الأساسية، مثل الحواف، الأشكال، والقوام السطحي، مما يوفّر تمثيلًا أوّليًا لبنية المشهد الداخلي. غير أن هذا الفهم المحلي، على الرغم من أهميته، يبقى محدودًا عند التعامل مع المشاهد المعقّدة التي تتطلب إدراك العلاقات العامة بين العناصر.هنا يبرز دور نماذج Transformers، التي تعالج هذه المحدودية من خلال إدراك السياق العام للصورة، وربط العناصر المتباعدة مكانيًا ضمن تمثيل واحد متكامل. يتيح هذا الإدراك السياقي فهم البنية العامة للمشهد الداخلي، مثل العلاقة بين الأرضية والجدران وتوزيع الأثاث ضمن المساحة المتاحة، وهو ما يعزّز دقة التمثيل البصري للمشهد.

بعد ذلك، يوفّر التقسيم الدلالي تمثيلًا تفصيليًا للمشهد على مستوى البكسل، من خلال إسناد تسمية دلالية لكل نقطة في الصورة. ويُعدّ هذا التمثيل خطوة أساسية لتمكين التعديل البصري الدقيق، إذ يسمح بتحديد العناصر المستهدفة للتعديل دون التأثير على باقي مكوّنات المشهد.

وأخيرًا، يأتي نموذج Segment Anything Model ليكمّل هذا التسلسل من خلال تحسين الدقة الهندسية للأقنعة الناتجة عن التقسيم الدلالي. إذ يساهم في تصحيح الحواف وتعزيز التماسك الهندسي بين العناصر، مما يؤدي إلى نتائج أكثر واقعية وتناسقًا. وبهذا، يتكامل الفهم الدلالي مع الدقة الهندسية ضمن نظام واحد، يوازن بين إدراك المعنى البصري ودقة التمثيل المكاني.

## خلاصة الفصل:

قدّم هذا الفصل دراسة نظرية موسّعة للتقنيات المعتمدة في البحث، بدءًا من مفاهيم التعلّم العميق بوصفه الأساس العام لمعالجة الصور، مرورًا بالشبكات العصبية الالتفافية التي شكّلت نقطة الانطلاق في استخلاص السمات البصرية، ثم نماذج Transformers التي أحدثت تحولًا نوعيًا في إدراك السياق العام للمشاهد.

كما تناول الفصل مفهوم التقسيم الدلالي وأهميته في تمثيل المشاهد الداخلية على مستوى البكسل، مبيّنًا دوره الأساسي في تحديد العناصر المعمارية بدقة. واختُتم العرض النظري بنموذج Segment Anything Model، الذي أُدخل كحل تكمِيلي لتحسين الدقة الهندسية للأقنعة الناتجة، ومعالجة القيود المرتبطة بدقة الحواف في النماذج الدلالية التقليدية.

يؤكّد هذا العرض النظري أن دمج هذه التقنيات ضمن إطار واحد متكامل يوفّر أساسًا علميًا متينًا لتطوير نظام قادر على تحليل المشاهد الداخلية بدقة وواقعية. ويمهّد هذا الفصل للانتقال إلى الفصل التالي، الذي يتناول المنهجية العملية المقترحة، ويشرح كيفية توظيف هذه المفاهيم والنماذج ضمن نظام واحد لتحقيق أهداف البحث.

# **الفصل الرابع: المعطيات الأولية وطرق تحصيلها**



## مقدّمة:

تُعد المعطيات الأولية الركيزة الأساسية في بناء الأنظمة المعتمدة على تقنيات التعلّم العميق، إذ تؤثر طبيعتها وجودتها بشكل مباشر على دقة النتائج وقدرة النماذج على التعميم. وفي مجال تحليل المشاهد الداخلية، تكتسب المعطيات أهمية خاصة نظرًا لتعقيد هذه المشاهد، وتداخل العناصر المعمارية مع الأثاث، إضافة إلى تنوّع أنماط التصميم الداخلي واختلاف ظروف الإضاءة وزوايا التصوير.يهدف هذا الفصل إلى عرض المعطيات الأولية المعتمدة في المشروع، وشرح مصادرها، وآلية تحصيلها، إضافة إلى توضيح عمليات التنقية والتجهيز التي خضعت لها قبل استخدامها ضمن النظام المقترح.



## أنواع المعطيات الأولية المعتمدة في المشروع:

اعتمد المشروع على نوعين رئيسيين من المعطيات الأولية، يختلف كل منهما من حيث المصدر وطبيعة الاستخدام ضمن النظام. شمل النوع الأول معطيات موسومة تُستخدم في تدريب نماذج التقسيم الدلالي، بينما شمل النوع الثاني معطيات غير موسومة خُصصت لتقييم أداء النظام واختبار قدرته على التعميم في بيئات واقعية خارج نطاق معطيات التدريب.

## المعطيات الأولية الأولى: صور مشاهد داخلية موسومة:

تم الاعتماد في هذا المشروع على مجموعة معطيات مرجعية موسومة تُعرف باسم NYUDepth V2، والتي تتضمن حوالي 1450 صورة ملونة(RGB) لمشاهد داخلية سكنية، مرفقة بتسميات دلالية دقيقة على مستوى البكسل، إضافة إلى ملفات توصيف بصيغة MATLAB تحتوي على معلومات الأقنعة الدلالية.

تشمل هذه المعطيات أنواعًا متعددة من الغرف السكنية مثل غرف النوم، المطابخ، غرف المعيشة، والحمّامات، مع تمثيل واضح للعناصر المعمارية الأساسية كالجدران، الأرضيات، والأسقف. وقد تم استخدام الصور بصيغتها الأصلية دون إجراء أي عمليات تعديل أو تنظيف بصري عليها، وذلك للحفاظ على التطابق الكامل بينها وبين الأقنعة الدلالية المرافقة، وتجنّب إدخال أي تشويه قد يؤثر على دقة عملية التدريب.

اقتصرت عملية تجهيز هذه المعطيات على التعامل مع ملفات التسميات بصيغة MATLAB، حيث تم تنظيمها ومعالجتها بما يتوافق مع متطلبات الاستخدام ضمن بيئة العمل البرمجية، دون المساس بمحتوى الصور نفسها، وذلك لضمان سلامة العلاقة بين الصورة والقناع الدلالي.

## المعطيات الأولية الثانية: صور مشاهد داخلية واقعية غير موسومة:

إلى جانب المعطيات الموسومة، تم الاعتماد على مجموعة معطيات غير موسومة تُعرف باسم Places365-Standard، وهي مجموعة واسعة تضم عددًا كبيرًا من الصور التي تمثل مشاهد داخلية وخارجية متنوعة. ونظرًا لضخامة حجم هذه المجموعة وتنوّع محتواها، لم يكن من الممكن استخدامها بالكامل ضمن نطاق هذا المشروع.

في المرحلة الأولى، تم اختيار الصور التي تمثل مشاهد داخلية سكنية فقط، مع استبعاد الصور الخارجية أو غير المرتبطة ببيئة الغرف المنزلية. بعد ذلك، خضعت الصور المختارة لعملية فحص وتنقية يدوية، حيث تم حذف الصور غير المناسبة، مثل الصور التي تحتوي على تشويش بصري واضح، أو الصور التي لا تمثل غرفة داخلية بشكل متكامل، أو تلك التي يظهر فيها تداخل غير منطقي للعناصر داخل المشهد.

خلال عملية التنقية، لوحظ نقص في تمثيل بعض أنواع الغرف، وبشكل خاص فئة **المداخل** (Entrance)، وهي فئة مهمة ضمن سياق تحليل المشاهد الداخلية السكنية. لمعالجة هذا النقص، تم جمع صور إضافية لهذه الفئة وإضافتها إلى مجموعة المعطيات، بهدف تحقيق تنوّع وتمثيل أفضل لأنواع الغرف المختلفة.

بعد الانتهاء من عمليات الاختيار والتنقية والإضافة، تم التوصل إلى مجموعة معطيات نهائية غير موسومة تتألف من حوالي 1100 صورة، تعكس تنوّعًا واضحًا في أنماط التصميم الداخلي، الإضاءة، وزوايا التصوير، مما يجعلها مناسبة لاختبار قدرة النظام على التعميم والعمل في بيئات واقعية خارج نطاق معطيات التدريب.

## طريقة تحصيل المعطيات:

تم تحصيل المعطيات الموسومة من مجموعة بيانات مرجعية مستخدمة على نطاق واسع في الأبحاث الأكاديمية، وتم اختيارها لما توفره من تسميات دلالية دقيقة وموثوقة لعناصر المشاهد الداخلية. في المقابل، تطلّبت المعطيات غير الموسومة جهدًا إضافيًا في التحصيل نظرًا لحجمها الكبير وتنوّع محتواها، حيث لم تكن جميع الصور المتوفرة مناسبة للاستخدام ضمن نطاق هذا المشروع.

بناءً على ذلك، تم اتباع آلية منهجية في تحصيل المعطيات غير الموسومة، بدأت باختيار الملفات التي تتوافق مع طبيعة المشاهد الداخلية السكنية وأهداف النظام المقترح، مع استبعاد الصور التي لا تخدم غاية التحليل أو لا تمثل غرفًا داخلية بشكل واضح. ساهم هذا الأسلوب في ضمان توافق المعطيات المستخدمة مع متطلبات المشروع منذ المراحل الأولى.

## تنظيف المعطيات وتجهيزها:

اختلفت آلية تنظيف وتجهيز المعطيات باختلاف نوعها وطبيعتها. بالنسبة للمعطيات الموسومة، لم تخضع الصور لأي عمليات تنظيف أو تعديل بصري، وذلك حفاظًا على التطابق الكامل بينها وبين التسميات الدلالية المرافقة لها، وتجنّب إدخال أي تغييرات قد تؤثر على سلامة العلاقة بين الصورة والقناع الدلالي.

اقتصرت عمليات التجهيز في هذه الحالة على التعامل مع ملفات التسميات بصيغة MATLAB، حيث تم تنظيمها ومعالجتها بما يضمن جاهزيتها للاستخدام ضمن بيئة العمل البرمجية، دون المساس بمحتوى الصور نفسها. ساهم هذا الإجراء في الحفاظ على دقة البيانات وضمان استخدامها بشكل صحيح خلال مراحل التدريب.

أما المعطيات غير الموسومة، فقد تطلّبت عمليات تنظيف وتجهيز أكثر تعقيدًا نظرًا لحجمها الكبير وطبيعتها غير المهيكلة. شملت هذه العمليات فحص الصور المختارة واستبعاد الصور غير المناسبة، مثل الصور التي تحتوي على تشويش بصري واضح، أو تلك التي لا تمثل غرفة داخلية بشكل متكامل، أو التي يظهر فيها تداخل غير منطقي للعناصر داخل المشهد. أدّت هذه الخطوات إلى رفع جودة المعطيات وضمان ملاءمتها لمرحلة التقييم.

## استخدام المعطيات ضمن النظام:

استُخدمت المعطيات الموسومة في تدريب نماذج التقسيم الدلالي والتحقق من أدائها، نظرًا لاحتوائها على تسميات دلالية دقيقة تمكّن من تقييم النتائج بشكل موضوعي. في المقابل، خُصصت المعطيات غير الموسومة لاختبار النتائج البصرية وتقييم قدرة النظام على التعميم، وذلك من خلال تطبيق النظام على صور واقعية غير مستخدمة في التدريب.

أتاح هذا الفصل الواضح بين معطيات التدريب ومعطيات التقييم الحصول على تقييم أكثر واقعية لأداء النظام، وساعد في قياس مدى فعاليته عند التعامل مع مشاهد داخلية حقيقية قريبة من ظروف الاستخدام الفعلي.

## خلاصة الفصل:

استعرض هذا الفصل المعطيات الأولية المعتمدة في المشروع وطرق تحصيلها وتجهيزها، مع توضيح الفروق الجوهرية بين المعطيات الموسومة وغير الموسومة ودور كل منها ضمن النظام المقترح. كما بيّن الجهد المبذول في اختيار وتنقية المعطيات غير الموسومة، وأهمية الحفاظ على سلامة المعطيات الموسومة دون تعديل بصري.

أسهمت هذه المنهجية في إعداد معطيات ذات جودة مناسبة للتدريب والتقييم، مما دعم موثوقية النتائج وعزّز قدرة النظام على التعميم عند تطبيقه على مشاهد داخلية واقعية.

# **الفصل الخامس :المنهجية المقترحة**



## مقدّمة الفصل:

يعرض هذا الفصل المنهجية العملية المعتمدة في هذا البحث لتطوير نظام ذكي قادر على فهم المشاهد الداخلية وتحليلها بصريًا بدقة، اعتمادًا على مجموعة من نماذج التعلّم العميق المتقدمة القائمة على بنى المحوّلات البصرية (Transformers) . فبعد أن تناولت الفصول السابقة الخلفية النظرية والمعطيات الأولية المستخدمة، يركّز هذا الفصل على توضيح كيفية تحويل تلك المفاهيم النظرية إلى نظام تطبيقي متكامل يحقق أهداف المشروع.

تعتمد المنهجية المقترحة على مبدأ خط المعالجة المتسلسل، حيث تُقسَّم عملية تحليل الصورة إلى مراحل واضحة ومترابطة، تؤدي كل مرحلة فيها دورًا محددًا يهيّئ البيانات للمرحلة اللاحقة. ويهدف هذا الأسلوب إلى تجاوز القيود المرتبطة بالاعتماد على نموذج واحد فقط، من خلال الاستفادة من نقاط القوة في كل نموذج ومعالجة نقاط الضعف عبر التكامل بينها، بما يضمن تحقيق توازن بين الدقة الدلالية والدقة الهندسية للمخرجات.

تستند هذه المنهجية إلى دمج ثلاثة نماذج رئيسية تعمل ضمن إطار واحد، هي: نموذج Swin Transformer Tiny لتصنيف نوع الغرفة بهدف توفير فهم سياقي أولي للمشهد، ونموذج SegFormer لإجراء التقسيم الدلالي على مستوى البكسل، ونموذج Segment Anything Model (SAM) لتحسين الدقة الهندسية للأقنعة الناتجة. ويُوظَّف هذا الدمج بطريقة مدروسة تضمن إنتاج مخرجات دقيقة وقابلة للاستخدام العملي، تمهيدًا لتطبيق تعديلات بصرية موجّهة على عناصر المشهد الداخلي ضمن سياق التصميم الداخلي.

## بنية النظام العامة:

يعتمد النظام المقترح على بنية متعددة المراحل تبدأ بإدخال صورة لمشهد داخلي خام، وتنتهي بإنتاج خرائط تقسيم وأقنعة دقيقة قابلة للاستخدام العملي في تطبيقات التعديل البصري. وقد تم تصميم هذه البنية بأسلوب يضمن انتقال البيانات بشكل منظم وسلس بين المراحل المختلفة، مع الحفاظ على الاتساق الدلالي للمحتوى والدقة الهندسية للمخرجات في كل مرحلة.

تعمل بنية النظام وفق تسلسل هرمي واضح، يمكن تلخيص مراحله على النحو الآتي:

• إدخال صورة مشهد داخلي واحدة تمثل المدخل الأساسي للنظام.

• تصنيف نوع الغرفة باستخدام نموذج تصنيف سياقي لتوفير فهم عام لهوية المشهد.

• إجراء التقسيم الدلالي على مستوى البكسل باستخدام نموذج SegFormer لاستخلاص المكونات الأساسية للمشهد.

• تحسين نتائج التقسيم الدلالي باستخدام نموذج Segment Anything Model للحصول على أقنعة ذات حدود هندسية أكثر دقة.

• دمج المخرجات الناتجة من جميع المراحل لإنتاج النتيجة النهائية القابلة للاستخدام.

يمثّل هذا التسلسل إطارًا منهجيًا مدروسًا، حيث تؤدي كل مرحلة دورًا داعمًا ومكمّلًا للمرحلة اللاحقة. فمرحلة تصنيف نوع الغرفة توفّر السياق العام الذي يوجّه تفسير نتائج التقسيم الدلالي، في حين يقدّم التقسيم الدلالي تمثيلًا تفصيليًا لمكوّنات المشهد يُستخدم لاحقًا كمدخل لتحسين الأقنعة. ويساهم هذا التكامل بين المراحل في رفع دقة النتائج النهائية، وتعزيز استقرار النظام عند التعامل مع صور واقعية متنوّعة وغير متجانسة.

## نموذج تصنيف نوع الغرفة:



### **دور النموذج ضمن النظام:**

يمثّل نموذج تصنيف نوع الغرفة المرحلة الأولى في خط المعالجة المقترح، ويهدف إلى تحديد نوع البيئة الداخلية التي تحتويها الصورة المدخلة، مثل غرفة نوم، غرفة معيشة، مطبخ، أو غيرها من الأنماط الشائعة. ويُعد هذا التحديد خطوة أساسية في فهم المشهد الداخلي، نظرًا لاختلاف توزيع العناصر المعمارية والأثاثية وطبيعة الاستخدام بين أنواع الغرف المختلفة.

يوفّر هذا النموذج فهمًا سياقيًا عامًا لهوية المشهد، يُستخدم لاحقًا لتوجيه تفسير نتائج التقسيم الدلالي ضمن سياقها الصحيح. فعلى سبيل المثال، يختلف تفسير وجود عناصر معينة مثل الخزائن أو الأسطح الواسعة باختلاف نوع الغرفة. ويساهم هذا الفهم المسبق في تقليل الالتباس الدلالي بين العناصر المتشابهة بصريًا ولكن المختلفة وظيفيًا، مما يعزّز موثوقية النظام عند التعامل مع مشاهد داخلية متنوّعة وغير متجانسة.



### **آلية عمل النموذج:**

يعتمد نموذج تصنيف نوع الغرفة في هذا المشروع على بنية Swin Transformer Tiny (swin\_tiny\_patch4\_window7\_224)، وهي بنية قائمة على المحوّلات البصرية صُمّمت لمعالجة الصور بكفاءة مع الحفاظ على تمثيل سياقي غني. تبدأ آلية العمل بتغيير حجم الصورة المدخلة إلى أبعاد ثابتة مقدارها 224×224، ثم تقسيمها إلى رقع صغيرة بحجم 4×4 بكسل، تُحوَّل إلى تمثيلات رقمية قابلة للمعالجة.

تُعالج هذه التمثيلات باستخدام آلية الانتباه الذاتي (Self-Attention) ضمن نوافذ محلية بحجم 7×7، مع تطبيق مفهوم النوافذ المتزاحة (Shifted Windows) بين الطبقات، مما يسمح بتبادل المعلومات بين الرقع المجاورة وتوسيع مجال الرؤية تدريجيًا دون زيادة كبيرة في الكلفة الحسابية. تتيح هذه الآلية للنموذج التقاط كلٍّ من السمات المحلية الدقيقة والعلاقات السياقية الأوسع داخل المشهد.

تمرّ الصورة عبر عدة مراحل هرمية داخل النموذج، حيث يتم دمج السمات البصرية من مستويات منخفضة تعبّر عن التفاصيل الأساسية، إلى مستويات أعلى تعبّر عن البنية العامة وترتيب العناصر داخل الغرفة. وفي المرحلة النهائية، تُحوَّل هذه السمات إلى مخرجات تصنيفية تمثل نوع الغرفة، مرفقة بدرجات ثقة تعبّر عن احتمالية انتماء الصورة لكل فئة مدروسة. وتُستخدم هذه المخرجات لاحقًا كمدخل سياقي داعم للمراحل اللاحقة في النظام.



### **القيمة المضافة للنموذج:**

تكمن القيمة المضافة لنموذج تصنيف نوع الغرفة في كونه يشكّل الأساس السياقي الذي تُبنى عليه بقية مراحل النظام المقترح. إذ يساهم هذا النموذج في:

• توفير إطار دلالي عام يوضّح هوية المشهد الداخلي قبل الدخول في التحليل التفصيلي.

• دعم نتائج التقسيم الدلالي من خلال وضعها ضمن سياقها الوظيفي الصحيح.

• تقليل الغموض الدلالي بين العناصر المتشابهة بصريًا عند اختلاف نوع الغرفة.

• تحسين استقرار النظام وموثوقية مخرجاته عند التعامل مع صور واقعية متنوّعة.

وبذلك، لا يُنظر إلى نموذج تصنيف الغرفة كمرحلة مستقلة، بل كعنصر موجِّه وأساسي يساهم في تحسين جودة خط المعالجة بأكمله.

## نموذج SegFormer للتقسيم الدلالي:



### **مبررات اختيار النموذج:**

يُعدّ SegFormer من النماذج الحديثة والمتقدمة في مجال التقسيم الدلالي، ويعتمد على بنية قائمة على المحوّلات البصرية (Transformers) صُمّمت خصيصًا لمعالجة الصور بكفاءة عالية مع الحفاظ على الفهم السياقي العام. ويتميّز هذا النموذج بقدرته على تجاوز القيود المرتبطة بنماذج التقسيم التقليدية القائمة على الشبكات العصبية الالتفافية، والتي تعتمد بشكل أساسي على المعالجة المحلية وقد تعاني من ضعف في تمثيل العلاقات بعيدة المدى داخل الصورة.

تم اختيار SegFormer في هذا المشروع نظرًا لقدرته على الدمج المتوازن بين السمات المحلية والسياق العام للمشهد، وهو عامل بالغ الأهمية عند تحليل المشاهد الداخلية التي تتسم بتداخل العناصر المعمارية وتشابه خصائصها البصرية، مثل الجدران والأسقف والأرضيات. كما يوفّر النموذج كفاءة حسابية مناسبة مقارنة بنماذج تقسيم أخرى ذات بنى أثقل، مما يجعله ملائمًا للتكامل ضمن نظام متعدد المراحل يهدف إلى إنتاج نتائج عملية قابلة للاستخدام.

إضافة إلى ذلك، يتميّز SegFormer بمرونته في التعامل مع الصور ذات الدقة المختلفة دون الحاجة إلى إعادة تصميم البنية أو تعديل جوهري في المعالجة، وهو ما يدعم استقرار النظام المقترح عند التعامل مع صور واقعية متنوعة المصدر والجودة.



### **آلية عمل SegFormer:**

تعتمد آلية عمل نموذج SegFormer على بنية ترميز–فك ترميز (Encoder–Decoder) قائمة على المحوّلات البصرية. في مرحلة الترميز، تُقسَّم الصورة المدخلة إلى رقع صغيرة تُحوَّل إلى تمثيلات رقمية، ثم تُعالج باستخدام طبقات Transformer متعدّدة المستويات لاستخلاص سمات بصرية تعبّر عن التفاصيل المحلية والعلاقات السياقية واسعة النطاق.

تتميّز بنية SegFormer بقدرتها على استخراج سمات متعددة المقاييس دون الاعتماد على عمليات التفاف متكررة أو إسقاطات مكانية معقّدة، مما يحدّ من فقدان المعلومات الدلالية أثناء المعالجة. وتُنتَج في هذه المرحلة عدة خرائط سمات تمثل مستويات مختلفة من التجريد البصري، بدءًا من التفاصيل الدقيقة وصولًا إلى التمثيل العام لبنية المشهد.

في مرحلة فك الترميز، يتم دمج السمات المستخرجة من المستويات المختلفة بطريقة مباشرة وفعّالة، ثم إعادة إسقاطها إلى أبعاد الصورة الأصلية لإنتاج خريطة تقسيم دلالي كاملة. يتيح هذا الأسلوب تصنيف كل بكسل في الصورة وفق الفئة التي ينتمي إليها، مثل الجدران، الأرضيات، أو الأسقف، مع الحفاظ على الاتساق الدلالي العام للمشهد.



### **دور SegFormer في المشروع:**

يمثّل نموذج SegFormer العنصر المركزي في فهم المشهد الداخلي على المستوى الدلالي، إذ يوفّر تمثيلًا تفصيليًا لمكوّنات الصورة يُستخدم كأساس لبقية مراحل النظام. وتُعد خرائط التقسيم الناتجة عن هذا النموذج المرجع الأساسي لتحديد مواقع الأسطح والعناصر المعمارية التي سيتم التعامل معها لاحقًا في مراحل التعديل البصري.

ورغم الدقة الدلالية العالية التي يوفّرها SegFormer، إلا أن الأقنعة الناتجة قد تعاني في بعض الحالات من عدم دقة الحدود الهندسية بين العناصر المتجاورة، خاصة في المناطق التي تتقارب فيها الخصائص اللونية أو الإضاءة. ويُعد هذا القيد متوقعًا في نماذج التقسيم الدلالي التي تركّز على التصنيف على مستوى البكسل أكثر من التركيز على دقة الحواف.

لذلك، لا يتم اعتماد مخرجات SegFormer بشكل نهائي، بل تُستخدم كمدخل أولي لمرحلة تحسين لاحقة باستخدام نموذج Segment Anything Model (SAM)، بهدف رفع دقة الحدود الهندسية للأقنعة وضمان جاهزيتها للاستخدام العملي في تطبيقات التعديل البصري الموجّه.

## نموذج Segment Anything Model (SAM):



### **دافع استخدام النموذج:**

يُستخدم نموذج Segment Anything Model (SAM) في هذا المشروع كمرحلة تحسين لاحقة لنتائج التقسيم الدلالي، نظرًا لقدراته المتقدّمة في تحديد الحدود الهندسية الدقيقة للعناصر داخل الصورة. وعلى عكس نماذج التقسيم الدلالي التقليدية التي تركّز بشكل أساسي على تصنيف البكسلات، يتميّز SAM بتركيزه على شكل العنصر وحدوده الهندسية، مما يجعله مناسبًا لمعالجة المشكلات المرتبطة بعدم دقة الحواف.

يعتمد هذا النموذج على تدريب واسع النطاق على مجموعات بيانات ضخمة ومتنوعة، مما يمنحه قدرة عالية على التعميم والتعامل مع أنواع مختلفة من الصور دون الحاجة إلى إعادة تدريب أو ضبط مخصّص لكل سيناريو استخدام. ويُعد هذا الأمر بالغ الأهمية في سياق المشروع، حيث يتم التعامل مع صور واقعية داخلية تختلف في الإضاءة، المواد، وأنماط التصميم.



### **آلية توظيف SAM ضمن النظام:**

يتم توظيف نموذج SAM ضمن النظام المقترح باعتباره مرحلة تحسين موجّهة لنتائج التقسيم الدلالي الناتجة عن نموذج SegFormer. حيث يستقبل SAM الصورة الأصلية إلى جانب الأقنعة الأولية المستخرجة من مرحلة التقسيم الدلالي، ويُستخدم هذا الدمج لتوجيه النموذج نحو المناطق المستهدفة داخل الصورة بدل الاعتماد على تقسيم عشوائي أو غير مقيّد.

تعتمد آلية التوجيه على استخدام مخرجات SegFormer لتحديد مناطق الاهتمام، مثل الجدران أو الأرضيات، ثم يعمل SAM على إعادة تحليل هذه المناطق بالتركيز على الشكل الهندسي وحدود العناصر. يسمح هذا الأسلوب بتوليد أقنعة أكثر دقة واستمرارية، مع تحسين واضح في فصل الحدود بين العناصر المتجاورة، دون التأثير على بقية أجزاء المشهد.

ويُسهم هذا التكامل بين النموذجين في الاستفادة من الدقة الدلالية التي يوفّرها SegFormer، والدقة الهندسية التي يتميّز بها SAM، ضمن خط معالجة واحد متكامل.



### **مخرجات النموذج ودورها:**

ينتج عن استخدام نموذج SAM أقنعة ذات جودة هندسية عالية تتميّز بحدود واضحة واستمرارية أفضل مقارنة بالأقنعة الدلالية الأولية. وتُعد هذه الأقنعة مناسبة للاستخدام المباشر في التطبيقات العملية التي تتطلب دقة عالية في تحديد مناطق التعديل البصري، مثل تطبيق الخامات أو الألوان على أسطح محددة داخل المشهد.

يساهم هذا التحسين في تقليل الأخطاء البصرية الناتجة عن تداخل الأقنعة أو عدم انتظام الحواف، كما يرفع مستوى الواقعية في النتائج النهائية، ويضمن أن تكون التعديلات البصرية محصورة بدقة ضمن المناطق المستهدفة فقط، دون التأثير على العناصر الأخرى في الصورة.

وبذلك، يشكّل نموذج SAM الحلقة الأخيرة في مرحلة تحليل المشهد، ويمهّد للانتقال من التمثيل التحليلي إلى التطبيق العملي للتعديل البصري ضمن النظام المقترح.

## تكامل النماذج وآلية إنتاج النتيجة النهائية:

تعمل النماذج الثلاثة المعتمدة في هذا المشروع ضمن إطار تكاملي ومتسلسل، حيث تم تصميم خط المعالجة بحيث تستفيد كل مرحلة من مخرجات المرحلة السابقة، مع معالجة القيود المرتبطة بها. يبدأ هذا التكامل بنموذج تصنيف نوع الغرفة، الذي يوفّر فهمًا سياقيًا عامًا لهوية المشهد الداخلي، ويشكّل الأساس الدلالي الذي تُفسَّر على ضوئه نتائج التحليل اللاحقة.

اعتمادًا على هذا السياق العام، يتم تمرير الصورة إلى نموذج SegFormer، الذي يتولى تحليل المشهد على مستوى البكسل وإنتاج خرائط تقسيم دلالي تفصيلية تحدد مواقع العناصر المعمارية الأساسية داخل الغرفة. وتُستخدم هذه الخرائط كتمثيل دلالي أولي لمكوّنات المشهد، إلا أنها قد تحتوي في بعض الحالات على حدود غير دقيقة بين العناصر المتجاورة.

لمعالجة هذه الإشكالية، يتم إدخال مخرجات SegFormer إلى نموذج Segment Anything Model، الذي يعمل على تحسين الدقة الهندسية للأقنعة من خلال إعادة تحليل حدود العناصر وأشكالها. يسمح هذا التسلسل بتحويل التمثيل الدلالي الأولي إلى أقنعة دقيقة هندسيًا وقابلة للاستخدام العملي.

ينتج عن هذا التكامل نظام متماسك قادر على تحويل صورة واحدة لمشهد داخلي إلى مجموعة من المخرجات الدقيقة التي تجمع بين الدقة الدلالية والدقة الهندسية. ويُعد هذا الأسلوب أكثر فاعلية من الاعتماد على نموذج واحد فقط، إذ يتيح الاستفادة من نقاط القوة لكل نموذج ومعالجة نقاط الضعف عبر الدمج المنهجي بينها، مما يوفّر أساسًا قويًا للتطبيقات العملية في مجال التعديل البصري والتصميم الداخلي.



### **آلية التفاعل مع المستخدم وتطبيق التعديل البصري:**

بعد الانتهاء من مراحل تحليل المشهد الداخلي واستخراج الأقنعة الدقيقة للعناصر المعمارية باستخدام النماذج المعتمدة ضمن النظام المقترح، يتم إتاحة هذه المخرجات ضمن واجهة استخدام تفاعلية تهدف إلى تمكين المستخدم من الاستفادة العملية من نتائج التحليل البصري. تُصمَّم هذه الواجهة لتكون وسيطًا بين مخرجات النماذج الذكية والمستخدم النهائي، دون الحاجة إلى أي تدخل تقني مباشر من قبله.

تبدأ عملية التفاعل بإدخال المستخدم لصورة مشهد داخلي واحدة، والتي تمثل المدخل الأساسي للنظام. بعد معالجة الصورة داخليًا، تُتاح للمستخدم إمكانية اختيار السطح المراد تعديله، مثل الجدران أو الأرضيات، اعتمادًا على الأقنعة الدلالية المحسّنة الناتجة عن نموذج Segment Anything Model. ويُعد هذا الاختيار خطوة محورية، إذ يضمن أن يكون التعديل موجّهًا بدقة إلى العنصر المستهدف فقط.

عقب تحديد السطح، يتمكّن المستخدم من اختيار المادة أو اللون المراد تطبيقه، حيث يقوم النظام بربط هذا الاختيار مباشرة بالقناع الهندسي الموافق للعنصر المحدد. ويجري تطبيق التعديل البصري حصريًا ضمن حدود القناع الناتج، دون التأثير على بقية عناصر المشهد، مما يضمن عزلًا دقيقًا للمنطقة المستهدفة ويحافظ على سلامة البنية البصرية العامة للصورة.

تتميّز عملية التعديل البصري بكونها تعديلًا موضعيًا محافظًا على الخصائص الأصلية للمشهد، إذ يتم الأخذ بعين الاعتبار الإضاءة والظلال والانسجام اللوني العام، بحيث يظهر التعديل المطبق متكاملًا مع الصورة الأصلية من الناحية البصرية. ويساهم هذا الأسلوب في تعزيز واقعية النتائج النهائية، ويحدّ من ظهور تشوّهات أو تأثيرات غير مرغوبة قد تنتج عن التعديل الشامل للصورة.

وبذلك، يوفّر النظام المقترح آلية تفاعل واضحة وبسيطة تُمكّن المستخدم من الاستفادة من الأقنعة الدقيقة الناتجة عن النماذج الذكية، وتحويلها إلى تعديلات بصرية موجّهة وقابلة للتطبيق العملي، دون الإخلال بتناسق المشهد الداخلي أو التأثير على عناصره غير المستهدفة.



### **التعديل البصري الموجّه (Region-based Editing) :**

بعد الانتهاء من مراحل تحليل المشهد الداخلي واستخراج الأقنعة الدقيقة للأسطح المستهدفة، يتم تطبيق مرحلة التعديل البصري الموجّه، والتي تُعد من أهم مراحل النظام المقترح. لا يعتمد هذا التعديل على تغيير الصورة كاملة، بل يتم تطبيقه حصريًا داخل حدود القناع الناتج عن نموذج Segment Anything Model، وبأسلوب بكسل-بكسل (Pixel-wise Editing).

يهدف هذا الأسلوب إلى ضمان أن يكون التعديل محصورًا بدقة ضمن المنطقة المستهدفة فقط، مع الحفاظ على الخصائص البصرية الأصلية للمشهد، مثل الإضاءة والظلال والملمس العام للصورة. ويساهم ذلك في إنتاج نتائج واقعية تتكامل بصريًا مع الصورة الأصلية دون ظهور تشوّهات أو تأثيرات مصطنعة.

تعديل الجدرانWalls)):

عند تعديل الجدران، يتم اعتماد أساليب دمج واعية بالمحتوى البصري، سواء في حالة تطبيق لون جديد أو مادة زخرفية.

في حالة تغيير اللون، يتم العمل ضمن فضاء لوني مناسب يسمح بفصل الإضاءة عن اللون، بحيث يتم تعديل القناة اللونية مع الحفاظ على قناة الإضاءة الأصلية، مما يضمن استمرار الظلال الطبيعية وعدم فقدان العمق البصري للجدار.

أما في حالة تطبيق مادة زخرفية، فيتم استخدام خامات قابلة للتكرار (Tileable Textures)، مع ضبط معدل التكرار بما يتناسب مع أبعاد الجدار داخل الصورة. ويتم إجراء عملية مزج إحصائي بين الخامة الجديدة والألوان الأصلية للجدار، بهدف تحقيق انسجام بصري وتقليل التباين غير المرغوب فيه.

تعديل الأرضيات Floors)):

تُعد عملية تعديل الأرضيات أكثر تعقيدًا من تعديل الجدران، نظرًا لتأثير المنظور واختلاف مقياس العناصر داخل الصورة. تبدأ العملية بتقدير عرض الأرضية اعتمادًا على القناع المستخرج، ثم تقدير الحجم التقريبي للبلاطات أو وحدات الخامة بوحدة البكسل.

بعد ذلك، يتم توليد الخامة بمنظور بصري يحاكي امتداد الأرضية في العمق، من خلال تطبيق تحجيم متدرّج بين الجزء القريب والجزء البعيد من الصورة. ويتم استخدام مزج تدريجي (Feathering) عند حدود القناع لمنع ظهور حواف حادة أو مناطق شفافة غير مرغوب فيها، مما يضمن اندماج الخامة الجديدة بسلاسة مع المشهد الأصلي.

تسهم هذه الآلية في الحفاظ على التناسق البصري للأرضية المعدّلة، وتمنع فقدان الواقعية الناتج عن الإسقاط المباشر أو التكرار غير المراعي للمنظور.



## آلية التقييم والمعايير المعتمدة:

تُعد مقاييس تقييم الأداء عنصرًا أساسيًا في تحليل فعالية النماذج المعتمدة في هذا المشروع، إذ تتيح قياس جودة النتائج بصورة كمية وموضوعية. ونظرًا لاختلاف طبيعة المهام بين تصنيف نوع الغرفة والتقسيم الدلالي للعناصر الداخلية، تم اعتماد مقاييس مختلفة تتوافق مع كل نوع من هذه المهام، وذلك وفقًا لما هو معتمد في الأدبيات العلمية.



### **مقياس F1-score1 Macro) وWeighted):**

يُعد مقياس **F1-score** من أهم معايير تقييم الأداء في مسائل **التصنيف متعدد الفئات (Multi-class Classification)**، وقد تم اعتماده كمقياس رئيسي في نموذج **تصنيف نوع الغرفة**. يجمع هذا المقياس بين **الدقة (Precision)** و**الاسترجاع (Recall)** في قيمة واحدة، مما يجعله أكثر تعبيرًا عن الأداء الحقيقي للنموذج، خاصة في حالات عدم توازن البيانات بين الفئات.

الصيغة الرياضية:

حيث تُحسب كل من الدقة والاسترجاع كما يلي:

ويتم اعتماد نسختين من هذا المقياس:

* : Macro F1-score يُحسب مقياس F1 لكل فئة على حدة ثم يؤخذ المتوسط الحسابي، مما يمنح جميع الفئات وزنًا متساويًا بغض النظر عن حجمها.
* : Weighted F1-score يُحسب المتوسط مع الأخذ بعين الاعتبار عدد العينات في كل فئة، مما يعكس التوزيع الحقيقي للبيانات.

تكمن أهمية هذا المقياس في كونه مناسبًا للتصنيف متعدد الفئات، وأقل تأثرًا بعدم توازن البيانات مقارنة بمقياس الدقة الكلية (Accuracy).

تم استخدام هذا المقياس فعليًا أثناء التدريب والتقييم واختيار النموذج الأفضل.



### **مصفوفة الالتباس (Confusion Matrix):**

تُستخدم مصفوفة الالتباس لتحليل أداء نموذج التصنيف بشكل تفصيلي، حيث توضح عدد العينات التي تم تصنيفها بشكل صحيح أو خاطئ لكل فئة. تمثل الصفوف القيم الحقيقية True) (Labels، بينما تمثل الأعمدة القيم المتوقعة (Predicted Labels).

تساعد هذه المصفوفة في:

* تحليل الأخطاء الشائعة بين الفئات المختلفة.
* فهم حالات الالتباس بين الأصناف المتقاربة بصريًا.
* تفسير أداء النموذج بصورة بصرية وعددية.

تم حساب مصفوفة الالتباس ورسمها وحفظها ضمن مخرجات النموذج.



### **تقرير التصنيف (Classification Report):**

يوفّر تقرير التصنيف ملخصًا شاملًا لأداء نموذج التصنيف، حيث يعرض القيم الخاصة بكل فئة على حدة، ويتضمن المقاييس التالية:

* Precision
* Recall
* F1-score
* Support (عدد العينات (

يسمح هذا التقرير بتقييم الأداء التفصيلي للنموذج على مستوى كل فئة، بدلًا من الاعتماد على مقياس إجمالي واحد قد يخفي ضعف الأداء في بعض الفئات.



### **مقياس Mean Intersection over Union (mIoU):**

يُعد مقياس Mean Intersection over Union (mIoU) المعيار الأساسي لتقييم أداء نماذج التقسيم الدلالي (Semantic Segmentation)، وقد تم استخدامه في نموذج تقسيم عناصر الغرفة الأساسية مثل الجدران والأرضيات والأسقف.

يقيس هذا المقياس نسبة التداخل بين القناع الحقيقي (Ground Truth) والقناع الناتج عن النموذج (Prediction)، ويُحسب لكل فئة على حدة كما يلي:

ثم يتم حساب المتوسط عبر جميع الفئات:

حيث C هو عدد الفئات.

تكمن أهمية هذا المقياس في:

* تقييم جودة الأقنعة بدقة عالية.
* معاقبة التداخلات الخاطئة بين الفئات.
* كونه المعيار الأكثر اعتمادًا في أبحاث التقسيم الدلالي.



### **مقياس** **Mean Accuracy:**

يمثل Mean Accuracy متوسط دقة تصنيف البكسلات عبر جميع الفئات، حيث يتم حساب نسبة البكسلات المصنفة بشكل صحيح لكل فئة ثم أخذ المتوسط الحسابي.

يُستخدم هذا المقياس كمؤشر داعم إلى جانب mIoU، إلا أنه أقل دقة في تقييم جودة الأقنعة، كونه لا يعاقب التداخلات الخاطئة بنفس شدة mIoU.

## مبررات اختيار المنهجية:

تم اعتماد المنهجية المقترحة في هذا المشروع استنادًا إلى مجموعة من الاعتبارات العلمية والتطبيقية، أبرزها الحاجة إلى تحقيق توازن بين الدقة الدلالية والدقة الهندسية عند تحليل المشاهد الداخلية. إذ أظهرت الدراسات السابقة أن الاعتماد على نموذج واحد غالبًا ما يؤدي إلى نتائج غير مستقرة عند التعامل مع صور واقعية تتسم بتنوّع الإضاءة، المواد، وزوايا التصوير.

كما جاء اختيار منهجية متعددة المراحل بهدف تجاوز محدودية النماذج الفردية، من خلال الاستفادة من نقاط القوة التي يوفّرها كل نموذج على حدة، ومعالجة نقاط الضعف عبر التكامل المنهجي بينها. فالتصنيف السياقي يساهم في فهم هوية المشهد، بينما يوفّر التقسيم الدلالي تمثيلًا تفصيليًا لمكوّناته، ويعمل تحسين الأقنعة على رفع جودة الحدود الهندسية للمخرجات.

إضافة إلى ذلك، راعت هذه المنهجية القدرة على التعامل مع صور واقعية خارج بيئات التدريب، دون الحاجة إلى ضبط مخصّص لكل حالة استخدام، وهو ما يُعد عاملًا أساسيًا في التطبيقات العملية. ويسهم هذا الأسلوب في إنتاج مخرجات دقيقة وقابلة للتطبيق في مجال التصميم الداخلي، مع الحفاظ على واقعية التعديلات البصرية واستقرار أداء النظام.

## خلاصة الفصل:

قدّم هذا الفصل عرضًا منهجيًا مفصّلًا للآلية المقترحة في هذا المشروع، موضحًا بنية النظام العامة وتسلسل مراحله، ودور كل نموذج ضمن خط المعالجة، بدءًا من تصنيف نوع الغرفة، مرورًا بالتقسيم الدلالي وتحسين الأقنعة، وصولًا إلى مرحلة التعديل البصري الموجّه. كما تم توضيح آلية تكامل النماذج المختلفة لإنتاج مخرجات تجمع بين الدقة الدلالية والدقة الهندسية، وقابلة للاستخدام العملي.

يمهّد هذا العرض المنهجي للانتقال إلى الفصل التالي، حيث سيتم استعراض النتائج العملية للنظام المقترح وتحليل أدائه تجريبيًا، مع تقييم جودة المخرجات من الناحيتين الكمية والبصرية، ومدى تحقيقه لأهداف المشروع.

# **الفصل السادس: النتائج والتقييم**



## مقدّمة الفصل:

يهدف هذا الفصل إلى عرض وتحليل النتائج الناتجة عن تنفيذ النظام المقترح لتعديل المشاهد الداخلية اعتمادًا على دمج عدة نماذج من الذكاء الاصطناعي ضمن خط معالجة واحد متكامل. تم تصميم هذا النظام ليعالج المشهد الداخلي ابتداءً من فهم نوع الغرفة، مرورًا بتقسيم العناصر المعمارية الأساسية، وصولًا إلى تحسين الأقنعة وتطبيق التعديلات البصرية بشكل واقعي.

يركّز هذا الفصل على دراسة أداء كل نموذج من نماذج النظام بشكل مستقل، وذلك بهدف تقييم فعاليته وحدوده الخاصة، قبل الانتقال إلى تحليل الأداء الكلي للنظام الناتج عن تكامل هذه النماذج ضمن سلسلة معالجة واحدة. يسمح هذا الأسلوب بعزل أثر كل مرحلة على جودة المخرجات النهائية، ويُسهِم في تقديم فهم أعمق لكيفية مساهمة كل نموذج في تحسين النتائج.

يعتمد عرض النتائج في هذا الفصل على مزيج من التحليل الكمي والتحليل النوعي. حيث يتم استخدام المقاييس الإحصائية القياسية لتقييم أداء النماذج في المراحل التي تتوفر فيها بيانات موسومة، في حين يتم الاعتماد على التحليل البصري النوعي لتقييم جودة المخرجات في المراحل التي يصعب توصيفها رقميًا، ولا سيما عند التعامل مع صور واقعية غير موسومة.

يهدف هذا النهج إلى تقديم تقييم عملي يعكس الأداء الحقيقي للنظام في سيناريوهات الاستخدام الواقعي، بدلًا من الاكتفاء بتقييم نظري أو رقمي مجرّد، مما يعزز من موثوقية النتائج المستخلصة في هذا الفصل.

## بيئة التنفيذ ومنهجية التقييم:

تم تنفيذ جميع التجارب الخاصة بالنظام المقترح ضمن بيئة برمجية موحّدة تعتمد على لغة Python، باستخدام مجموعة من مكتبات التعلم العميق والرؤية الحاسوبية الشائعة في تطوير النماذج المعتمدة على الشبكات العصبية العميقة. ساهم توحيد بيئة التنفيذ في ضمان ثبات النتائج وقابليتها للمقارنة بين مختلف مراحل النظام. اعتمدت منهجية التقييم في هذا الفصل على مستويين متكاملين، وذلك بما يتناسب مع طبيعة كل نموذج والبيانات المتاحة له:

**أولًا: التقييم الكمي**

تم استخدام مجموعة من مقاييس الأداء الإحصائية القياسية، مثل Accuracy وPrecision وRecall وF1-score، لتقييم أداء النماذج في المراحل التي تتوفر فيها بيانات موسومة، ولا سيما في مرحلة تصنيف نوع الغرفة. تتيح هذه المقاييس قياس دقة النموذج وقدرته على التمييز بين الفئات المختلفة، إضافةً إلى تقييم التوازن بين الحساسية والدقة في التنبؤات.

**ثانيًا: التقييم النوعي البصري**

في المراحل التي يكون فيها التقييم الرقمي محدودًا أو غير ممكن، مثل مراحل التقسيم الدلالي، تحسين الأقنعة، والتعديل البصري، تم الاعتماد على التحليل البصري النوعي للمخرجات. يركّز هذا النوع من التقييم على دراسة جودة الأقنعة الناتجة، دقة الحدود بين العناصر، ومدى واقعية التعديلات البصرية من حيث الإضاءة، الظلال، وتجانس المواد مع المشهد الأصلي.

يسمح الجمع بين التقييم الكمي والتقييم النوعي بتقديم صورة شاملة عن أداء النظام، تجمع بين الدقة الإحصائية من جهة، والواقعية التطبيقية من جهة أخرى. ويُعد هذا الأسلوب مناسبًا لتقييم الأنظمة المعتمدة على معالجة الصور والمشاهد الداخلية، حيث لا يمكن الاعتماد على المقاييس الرقمية وحدها للحكم على جودة النتائج النهائية.

## نتائج نموذج تصنيف نوع الغرفة (Room Classification Model):



### **إعدادات النموذج وإجراءات التدريب:**

اعتمدت مرحلة تصنيف نوع الغرفة على استخدام نموذج Swin Transformer (Swin-Tiny) بوصفه معمارية حديثة وفعّالة في مهام تصنيف الصور، لما يتمتع به من قدرة عالية على التقاط العلاقات المكانية الهرمية داخل المشاهد الداخلية. تم توحيد حجم صور الإدخال إلى 224×224 بما يتوافق مع متطلبات النموذج ويضمن استقرار عملية التدريب.

أُجري تدريب النموذج باستخدام دفعات بيانات مصغّرة Mini-batches) ) بحجم 32 صورة، وذلك من خلال آلية تحميل البيانات .(DataLoader) تم تفعيل خلط عينات مجموعة التدريب بهدف تعزيز تعميم النموذج وتقليل تحيزه للترتيب المكاني للبيانات، في حين تم إيقاف خلط بيانات مجموعة التحقق لضمان تقييم ثابت ومتّسق للأداء خلال مراحل التدريب المختلفة.

ولمعالجة مشكلة عدم توازن عدد العينات بين فئات الغرف المختلفة، تم اعتماد أوزان للفئات Class Weights) ) محسوبة استنادًا إلى تكرار كل فئة ضمن بيانات التدريب. جرى دمج هذه الأوزان ضمن دالة الخسارة، بما يتيح للنموذج إعطاء أهمية أكبر للفئات الأقل تمثيلًا ويُسهم في تحسين التوازن العام في الأداء.

استُخدمت دالة الخسارة CrossEntropyLoss مع تطبيق تقنية Label Smoothing بقيمة 0.1، وذلك بهدف تقليل الثقة المفرطة في التنبؤات وتحسين قدرة النموذج على التعميم عند التعامل مع بيانات غير مرئية.

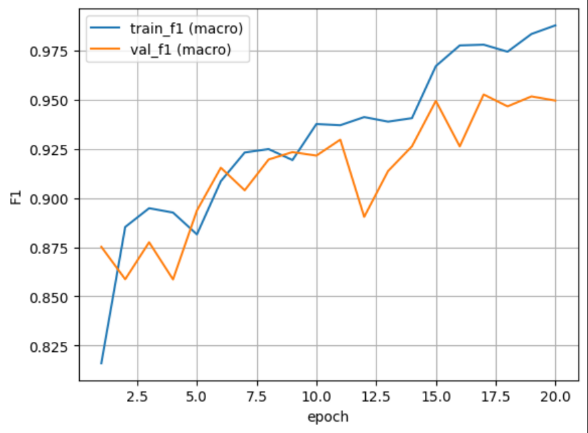
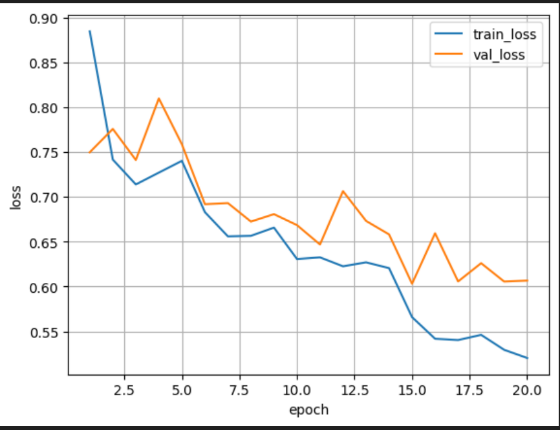
في مرحلة التحسين، تم اعتماد المُحسّن AdamW بمعدل تعلّم ابتدائي قدره 3×10⁻⁴، لما يوفره من استقرار في عملية التحديث وتنظيم فعّال للأوزان عبر آلية Weight Decay، وهو ما ينعكس إيجابًا على تقليل فرط التعلّم.

كما تم استخدام مجدول معدل التعلّم ReduceLROnPlateau لتنظيم قيمة معدل التعلّم ديناميكيًا أثناء التدريب، حيث يتم تخفيضه عند توقف تحسن أداء النموذج على مجموعة التحقق. تم ضبط معامل التخفيض (factor) بقيمة 0.5، مع patience = 2 ، بما يسمح بإعطاء النموذج فرصة كافية للتحسن قبل تقليل معدل التعلّم.

استمر تدريب النموذج لمدة أقصاها 20 حقبة تدريبية (Epochs)، مع تطبيق آلية الإيقاف المبكر (Early Stopping) استنادًا إلى تحسن قيمة F1-Score (macro) على مجموعة التحقق، وبقيمة patience = 5 لمحاولات عدم التحسن المتتالية. جرى حفظ أفضل نموذج تلقائيًا عند تحقيق أعلى أداء على مجموعة التحقق، لاستخدامه لاحقًا في مرحلة التقييم النهائي.



### **تحليل منحنيات التدريب والتحقق:**



**الشكل ( (6-1**

**الشكل (6-2)**

يوضح الشكل ( (6-1تطور قيم الخسارة لكل من مجموعة التدريب ومجموعة التحقق عبر الحِقَب التدريبية. يُلاحظ في المراحل الأولى من التدريب انخفاض حاد نسبيًا في قيمة Training Loss، ما يشير إلى قدرة النموذج على تعلّم السمات التمييزية الأساسية للمشاهد الداخلية بسرعة. في المقابل، يُظهر منحنى Validation Loss انخفاضًا عامًا مع وجود بعض التذبذبات المحدودة، خاصة في الحِقَب المتوسطة.

تعكس هذه التذبذبات سلوكًا طبيعيًا في نماذج التصنيف العميقة، وغالبًا ما يكون ناتجًا عن تنوع العينات داخل مجموعة التحقق واختلاف تعقيد المشاهد بين الفئات. وعلى الرغم من ذلك، فإن عدم وجود ارتفاع متزايد أو انفصال واضح بين منحنيي التدريب والتحقق يدل على أن النموذج لم يدخل في حالة فرط التعلّم Overfitting)) .

في المراحل المتقدمة من التدريب، يُلاحظ استمرار انخفاض قيمة Training Loss مقابل استقرار نسبي في Validation Loss، مما يشير إلى وصول النموذج إلى حالة تقارب (Convergence) مناسبة قبل نهاية عدد الحِقَب المحدد.

أما الشكل ( (6-1، فيعرض تطور قيمة F1-Score (Macro) لكل من مجموعة التدريب ومجموعة التحقق. تُظهر النتائج تحسنًا تدريجيًا واضحًا في أداء النموذج مع تقدم التدريب، حيث ترتفع قيمة F1-Score لكلا المجموعتين بشكل متوازن نسبيًا.

يُلاحظ أن قيمة F1-Score (Macro) على مجموعة التحقق تشهد بعض الانخفاض المؤقت في الحِقَب المتوسطة، قبل أن تعود إلى الارتفاع والاستقرار في الحِقَب اللاحقة. يعكس هذا السلوك حساسية النموذج لتغير توزيع العينات أثناء التدريب، دون أن يؤثر سلبًا على الأداء النهائي.

كما تُظهر الفجوة المحدودة بين منحنيي التدريب والتحقق في قيم F1-Score أن النموذج حافظ على توازن جيد بين القدرة على التعلّم من بيانات التدريب والقدرة على التعميم على بيانات غير مرئية. ويُعزى هذا الأداء المستقر إلى الاستراتيجيات المعتمدة أثناء التدريب، مثل استخدام أوزان الفئات، وتقنية Label Smoothing، وتنظيم معدل التعلّم، إضافةً إلى تطبيق آلية الإيقاف المبكر.

بناءً على تحليل منحنيات الخسارة ومقاييس الأداء، يمكن الاستنتاج أن النموذج وصل إلى أفضل أداء له في المراحل المتقدمة من التدريب، وأن اعتماد آلية الإيقاف المبكر أسهم في اختيار نموذج يحقق توازنًا مناسبًا بين الدقة والاستقرار دون الحاجة إلى استكمال التدريب حتى آخر الحِقَب الممكنة.



### **نتائج التقييم النهائي لنموذج تصنيف نوع الغرفة**

### **(Final Evaluation Results – Classification Report)**

بعد الانتهاء من تدريب نموذج تصنيف نوع الغرفة وتحليل سلوك التدريب والتحقق، تم إجراء التقييم النهائي للنموذج باستخدام مجموعة الاختبار غير المستخدمة في التدريب. يهدف هذا التقييم إلى قياس الأداء الحقيقي للنموذج وقدرته على التعميم على بيانات جديدة، وذلك بالاعتماد على مجموعة من مقاييس الأداء القياسية.

| **نوع الغرفة** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| Bathroom | 0.9829 | 0.9631 | 0.9729 | 298 |
| Bedroom | 0.9567 | 0.9493 | 0.9530 | 651 |
| Living Room | 0.8981 | 0.9153 | 0.9066 | 366 |
| Dining Room | 0.9566 | 0.9888 | 0.9725 | 357 |
| Entrance | 0.9812 | 0.9414 | 0.9609 | 222 |
| Home Office | 0.9219 | 0.9440 | 0.9328 | 375 |
| Kitchen | 0.9750 | 0.9613 | 0.9681 | 569 |
| **Macro Average** | **0.9532** | **0.9519** | **0.9524** | 2838 |
| **Weighted Average** | **0.9529** | **0.9524** | **0.9525** | 2838 |
| **Accuracy** | — | — | **0.9524** | 2838 |

**جدول (1-6): تقرير التصنيف النهائي**

تُظهر نتائج التقييم النهائي أن نموذج تصنيف نوع الغرفة حقق دقة كلية (Accuracy) بلغت 95.24%، وهو ما يعكس أداءً مرتفعًا وقدرة جيدة على التمييز بين مختلف فئات الغرف الداخلية.

سجّلت فئات مثل Bathroom وDining Room وKitchen قيمًا مرتفعة في مقاييس Precision وRecall وF1-score، مما يدل على قدرة النموذج على التعرّف بدقة على هذه الأنواع من الغرف. ويُعزى هذا الأداء المرتفع إلى الخصائص البصرية المميزة لهذه الفئات، مثل طبيعة التجهيزات، توزيع العناصر، والأنماط اللونية الواضحة التي تسهّل عملية التمييز بينها.

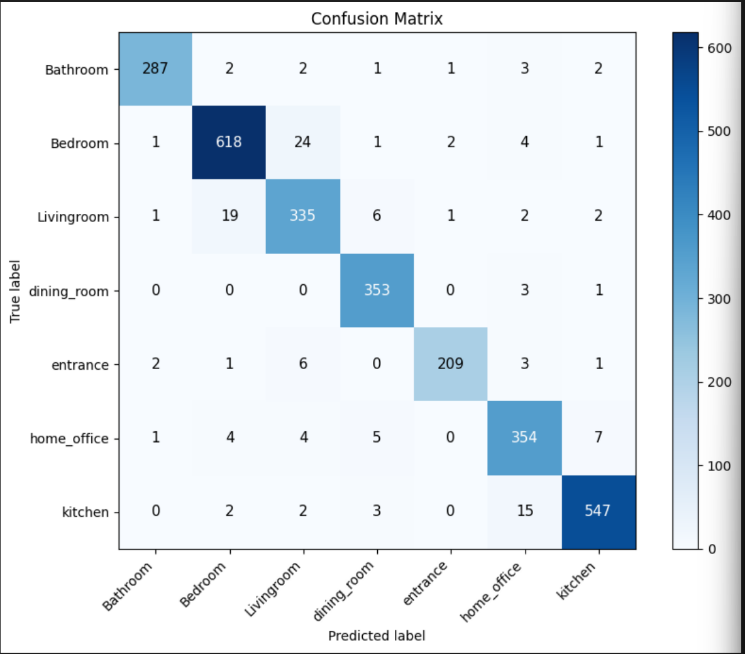
في المقابل، لوحظ انخفاض نسبي في أداء فئة Living Room مقارنةً ببقية الفئات، حيث سجلت أقل قيمة لـ. F1- score يُفسَّر هذا الانخفاض بالتشابه البصري الكبير بين غرف المعيشة وبعض الفئات الأخرى، مثل غرف النوم أو غرف الطعام، خاصة في الحالات التي تتداخل فيها أنماط الأثاث أو توزيع المساحات الداخلية.كما تُظهر القيم المتقاربة لكل من Macro Average وWeighted Average أن أداء النموذج متوازن نسبيًا عبر مختلف الفئات، وأنه لم يتأثر

بشكل كبير بعدم توازن عدد العينات بين الفئات المختلفة. ويعكس ذلك فعالية الاستراتيجيات المعتمدة أثناء التدريب، مثل استخدام أوزان الفئات، في تحسين التوازن العام للأداء.

بشكل عام، تؤكد نتائج التقييم النهائي أن نموذج تصنيف نوع الغرفة قادر على تقديم تنبؤات دقيقة وموثوقة، مما يجعله مرحلة تمهيدية فعالة تدعم المراحل اللاحقة في النظام، مثل التقسيم الدلالي وتحسين الأقنعة والتعديل البصري.



### **مصفوفة الالتباس (Confusion Matrix):**



**الشكل (6-3)**

يوضح الشكل (6-3) مصفوفة الالتباس الخاصة بنموذج تصنيف نوع الغرفة، والتي تبيّن العلاقة بين الفئات الحقيقية والفئات المتوقعة من قبل النموذج على مجموعة الاختبار. يُلاحظ من القيم المتمركزة على القطر الرئيسي للمصفوفة أن الغالبية العظمى من العينات قد تم تصنيفها بشكل صحيح، مما يعكس الأداء المرتفع للنموذج.

تُظهر فئات مثل Bathroom وDining Room وKitchen معدلات تصنيف صحيحة مرتفعة جدًا، حيث تتركز معظم العينات الخاصة بهذه الفئات على القطر الرئيسي دون وجود تداخل ملحوظ مع فئات أخرى. ويُعزى ذلك إلى الخصائص البصرية الواضحة لهذه الغرف، مثل طبيعة التجهيزات الداخلية وتوزيع العناصر المعمارية، مما يسهل عملية التمييز بينها.

في المقابل، يظهر بعض الالتباس المحدود بين فئة Living Room وفئات أخرى مثل Bedroom، وهو سلوك متوقع نظرًا للتشابه البصري بين هذه الأنواع من الغرف، خاصة في الحالات التي تتقارب فيها أنماط الأثاث أو المساحات المفتوحة. كما لوحظ تداخل بسيط بين Home Office وبعض الفئات المجاورة بصريًا، إلا أن هذا التداخل لم يؤثر بشكل كبير على الأداء الكلي للنموذج.

بشكل عام، تشير مصفوفة الالتباس إلى أن أخطاء التصنيف المتبقية تتركز في فئات ذات حدود بصرية غير واضحة، في حين يحافظ النموذج على دقة عالية في الفئات ذات السمات المميزة. وتؤكد هذه النتائج أن نموذج تصنيف نوع الغرفة يتمتع بقدرة جيدة على التمييز بين الفئات المختلفة، مع هامش خطأ محدود يمكن تفسيره بطبيعة البيانات نفسها.



### **مناقشة وخلاصة نموذج تصنيف نوع الغرفة:**

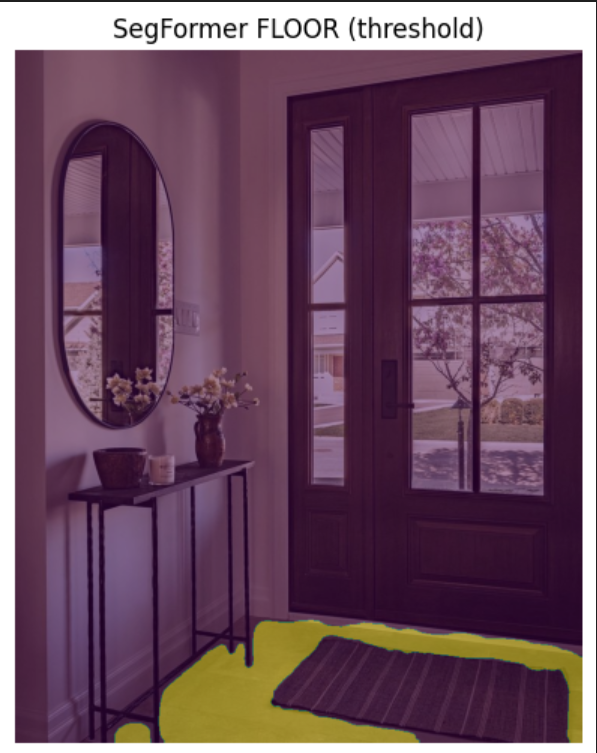
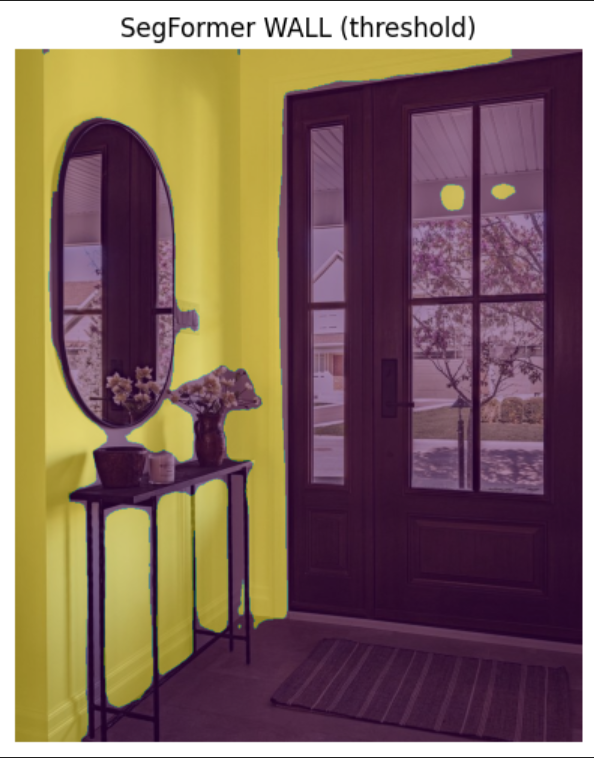
قدّم هذا القسم تحليلًا شاملًا لأداء نموذج تصنيف نوع الغرفة، بدءًا من إعدادات التدريب، مرورًا بسلوك النموذج أثناء عملية التعلم، وصولًا إلى التقييم النهائي باستخدام مقاييس الأداء القياسية ومصفوفة الالتباس. أظهرت النتائج أن النموذج قادر على تعلّم السمات البصرية المميِّزة للمشاهد الداخلية بكفاءة، مع تحقيق أداء مرتفع ومستقر عبر مختلف مراحل التدريب.

أوضحت منحنيات الخسارة ومقاييس F1-Score (Macro) أن عملية التدريب تمت بشكل متوازن دون مؤشرات واضحة على فرط التعلّم، كما بيّنت فعالية الاستراتيجيات المعتمدة مثل استخدام أوزان الفئات، وتنظيم معدل التعلّم، وآلية الإيقاف المبكر في تحسين قدرة النموذج على التعميم.

من ناحية التقييم النهائي، حقق النموذج دقة كلية مرتفعة، مع أداء قوي في معظم فئات الغرف، ولا سيما الفئات ذات الخصائص البصرية الواضحة مثل المطابخ والحمّامات وغرف الطعام. في المقابل، أظهر النموذج بعض الالتباس المحدود بين الفئات المتقاربة بصريًا، مثل غرف المعيشة وغرف النوم، وهو أمر متوقع في سياق تصنيف المشاهد الداخلية ويعود إلى طبيعة البيانات نفسها أكثر من كونه قصورًا في النموذج.

بشكل عام، تؤكد نتائج هذا القسم أن نموذج تصنيف نوع الغرفة يشكّل مرحلة تمهيدية موثوقة ضمن النظام المقترح، حيث يوفّر فهمًا أوليًا دقيقًا لطبيعة المشهد الداخلي. ويُعد هذا الفهم عنصرًا أساسيًا لدعم المراحل اللاحقة من النظام، مثل التقسيم الدلالي، تحسين الأقنعة، وتطبيق التعديلات البصرية بما يتناسب مع نوع الغرفة المكتشف.

## نتائج التقسيم الدلالي باستخدام: SegFormer



**الشكل6-4)): نتائج التقسيم الدلالي للجدران باستخدام SegFormer**

**الشكل(6-5): نتائج التقسيم الدلالي للأرضيات باستخدام SegFormer**

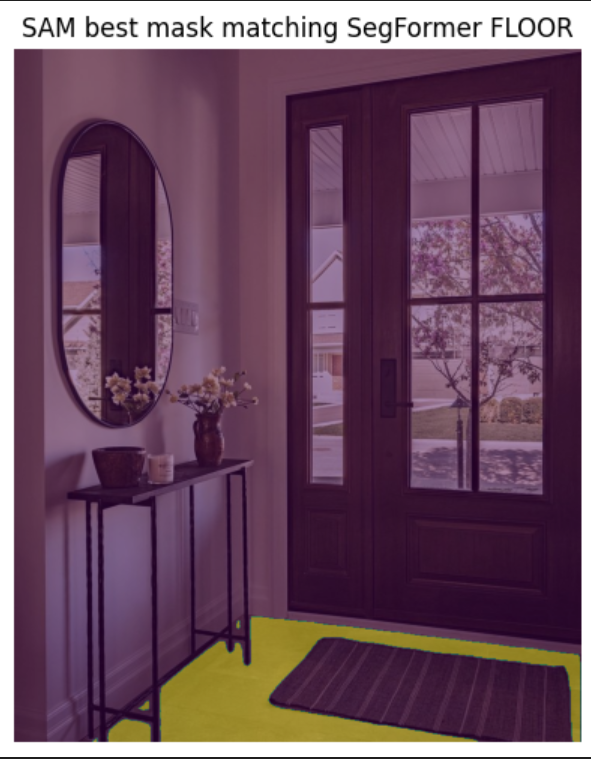
تُظهر الأشكال (6-4) و(6-5) نتائج التقسيم الدلالي المستخرجة باستخدام نموذج SegFormer لكل من الجدران والأرضيات داخل المشهد الداخلي. حيث تمكّن النموذج من تحديد الامتداد العام للأسطح المستهدفة بشكل واضح، مما يوفّر أساسًا مناسبًا لمراحل التعديل البصري اللاحقة.

بالنسبة للجدران، يبيّن القناع الناتج قدرة النموذج على تغطية معظم المساحات العمودية داخل المشهد، مع الحفاظ على الشكل العام للجدار. إلا أن بعض التداخلات المحدودة تظهر عند الحدود القريبة من عناصر أخرى مثل الأثاث أو الإكسسوارات، خاصة في المناطق التي يقل فيها التباين البصري.

أما في حالة الأرضيات، فقد أظهر النموذج قدرة جيدة على تحديد الامتداد الأفقي للأرضية، خصوصًا في المناطق المفتوحة. ومع ذلك، لوحظ وجود بعض الصعوبات عند حدود الأرضية، لا سيما في المناطق التي تحتوي على سجادات أو ظلال قوية، مما يؤدي أحيانًا إلى دمج أجزاء غير مستهدفة ضمن القناع.

بشكل عام، تؤكد هذه النتائج أن نموذج SegFormer يوفّر أقنعة دلالية أولية مناسبة لتمثيل البنية العامة للعناصر المعمارية داخل المشهد، إلا أن هذه الأقنعة تبقى تقريبية من حيث دقة الحدود. ويبرز ذلك الحاجة إلى مرحلة لاحقة لتحسين جودة الأقنعة قبل استخدامها في التعديل البصري، وهو ما يتم تحقيقه في المرحلة التالية باستخدام نموذج Segment Anything Model (SAM).

## تحسين الأقنعة باستخدام Segment Anything Model (SAM):



بعد الحصول على الأقنعة الدلالية الأولية الناتجة عن نموذج SegFormer، تم استخدام نموذج Segment Anything Model (SAM) بهدف تحسين جودة هذه الأقنعة قبل استخدامها في مرحلة التعديل البصري. تأتي هذه الخطوة لمعالجة القصور المرتبط بدقة الحدود والتداخل المحدود بين العناصر المتجاورة، والذي يُعد من الخصائص المتوقعة في مخرجات التقسيم الدلالي الأولي.

يعتمد نموذج SAM على آلية تقسيم مرنة تتيح إعادة تحديد حدود العناصر بدقة أعلى اعتمادًا على المحتوى البصري للصورة، دون الحاجة إلى تدريب مخصّص على فئات محددة. وبالاستفادة من الأقنعة الأولية المستخرجة من SegFormer، يتم توجيه النموذج نحو المناطق المستهدفة داخل المشهد، مما يسمح بتحسين الحواف وتقليل التسرب إلى العناصر غير المرغوب فيها.

تُظهر النتائج تحسنًا واضحًا في جودة الأقنعة من حيث نعومة الحدود ودقة الفصل بين العناصر المعمارية المختلفة، حيث أصبحت الحدود أكثر وضوحًا واستمرارية مقارنة بالأقنعة الأولية. كما اتّسمت الأقنعة الناتجة بدرجة أعلى من التماسك والاستقرار، مما يجعلها ملائمة للاستخدام المباشر في مراحل التعديل البصري اللاحقة، مثل تغيير المواد أو الألوان، دون التأثير على بقية مكوّنات المشهد.

بشكل عام، يساهم دمج نموذج SAM ضمن خط المعالجة المقترح في رفع جودة الأقنعة الدلالية وتحويلها من تقسيم تقريبي إلى أقنعة محسّنة ذات دقة هندسية أعلى، وهو ما يشكّل خطوة أساسية لضمان واقعية النتائج النهائية للنظام.

## نتائج التعديل البصري للعناصر المعمارية:

**الصورة الأصلية**

**التعديل الأول على الأرضية**



**التعديل الثاني على الأرضية**



**لتعديل الثاني**



**التعديل الأول**

**الصورة الأصلية**

تظهر الأمثلة الواردة في هذا القسم نتائج تطبيق الخامات والألوان المختلفة على العناصر المعمارية داخل المشهد الداخلي، من خلال مقارنة الصورة الأصلية بالمشاهد المعدّلة. في هذه المرحلة، تم تطبيق التعديلات البصرية اعتمادًا على الأقنعة المحسّنة الناتجة عن دمج نموذج SegFormer مع نموذج Segment Anything Model (SAM)، وذلك بهدف محاكاة تغييرات مادية واقعية مثل تغيير نوع الأرضية أو تشطيب الجدران، مع الحفاظ على الخصائص البصرية الأساسية للصورة الأصلية.

على الرغم من توفر أقنعة دلالية إضافية لبعض العناصر الأخرى، مثل السقف، إلا أنه لم يتم استخدامها في هذه المرحلة. ويعود ذلك إلى تركيز النظام الحالي على العناصر الأكثر تأثيرًا في المظهر البصري العام للمشهد الداخلي، والمتمثلة بالجدران والأرضيات. ومع ذلك، يتيح التصميم المرن للنظام إمكانية توسيع نطاق التعديلات مستقبلًا ليشمل عناصر إضافية عند الحاجة.

يوفّر النظام مرونة في آلية التعديل تبعًا لنوع العنصر المعماري، حيث يمكن للمستخدم عند التعامل مع الجدران اختيار تغيير لون الجدار أو تطبيق خامة زخرفية، بينما يقتصر التعديل على تطبيق الخامات فقط عند التعامل مع الأرضيات، بما يتوافق مع الاستخدام الواقعي لهذه العناصر في التصميم الداخلي.

أظهرت النتائج أن استخدام الأقنعة المحسّنة أتاح تطبيق الخامات أو الألوان بشكل موضعي ودقيق، حيث اقتصر تأثير التعديل على المناطق المستهدفة فقط، دون امتداده إلى العناصر المجاورة داخل المشهد. كما ساهمت دقة الحدود الناتجة عن تحسين الأقنعة في الحد من التداخل بين الخامات والعناصر الأخرى، مما عزّز وضوح النتائج واستقرارها بصريًا.

من ناحية الواقعية البصرية، حافظت الصور الناتجة بعد تطبيق الخامات أو الألوان على الإضاءة الطبيعية والظلال الأصلية للمشهد، الأمر الذي ساعد على دمج التعديلات الجديدة بشكل متجانس مع الصورة الأصلية. كما انسجمت المواد المطبّقة مع المنظور الهندسي للمشهد دون إحداث تشوّهات أو فقدان للتناسق البصري.

بشكل عام، تؤكد نتائج هذه المرحلة أن النظام المقترح قادر على تنفيذ تعديلات بصرية واقعية وموثوقة على العناصر المعمارية داخل المشاهد الداخلية، مع توفير مرونة للمستخدم في اختيار نوع التعديل المناسب لكل عنصر. ويبرز ذلك فعالية خط المعالجة المتكامل في دعم سيناريوهات تعديل المشاهد الداخلية بطريقة عملية وقابلة للتطبيق.

## نتائج الاختبار على صور واقعية غير موسومة:

في هذه المرحلة، تم اختبار النظام المقترح على صور واقعية غير موسومة لم تُستخدم خلال مراحل التدريب أو التحقق، وذلك بهدف تقييم قدرة النظام على التعميم والعمل بكفاءة في سيناريوهات قريبة من الاستخدام العملي الفعلي. ويُعد هذا النوع من الاختبارات ضروريًا للتحقق من موثوقية النظام خارج نطاق البيانات المعروفة مسبقًا.

أظهرت النتائج أن النظام قادر على معالجة الصور الواقعية بكفاءة، حيث نجح في تحديد العناصر المعمارية المستهدفة داخل المشهد، واستخراج الأقنعة المناسبة لها، ثم تطبيق الخامات أو الألوان بشكل متناسق مع البنية البصرية العامة للصورة. كما حافظ الأداء العام للنظام على مستوى جيد من الدقة والاستقرار، رغم اختلاف ظروف الإضاءة وتنوّع أنماط التصميم الداخلي في هذه الصور.

وبيّنت النتائج أيضًا أن دمج نماذج تصنيف نوع الغرفة، والتقسيم الدلالي، وتحسين الأقنعة باستخدام نموذج Segment Anything Model، وتطبيق الخامات ضمن خط معالجة واحد، مكّن النظام من التعامل مع صور غير مألوفة دون الحاجة إلى تدخل يدوي إضافي. ويؤكد ذلك قدرة النظام على التكيّف مع مشاهد داخلية متنوعة.

بشكل عام، تدعم نتائج الاختبار على الصور الواقعية غير الموسومة فعالية النظام المقترح، وتؤكد أنه لا يقتصر على تحقيق أداء جيد على بيانات التدريب فقط، بل يمتلك قدرة مناسبة على التعميم، وهو ما يُعد مؤشرًا مهمًا على جاهزيته للاستخدام العملي.

# **الفصل السابع: الاستنتاجات والعمل المستقبلي**



## خلاصة عامة:

قدّم هذا البحث نظامًا متكاملًا يعتمد على تقنيات الذكاء الاصطناعي لمعالجة وتعديل المشاهد الداخلية، من خلال دمج عدة نماذج ضمن خط معالجة واحد يشمل تصنيف نوع الغرفة، التقسيم الدلالي للعناصر المعمارية، تحسين الأقنعة، وتطبيق الخامات أو الألوان على الجدران والأرضيات. وقد أظهرت النتائج أن هذا التكامل يتيح تحقيق تعديلات بصرية موضعية مع الحفاظ على التناسق العام للمشهد.

بيّنت التجارب أن استخدام نماذج متقدمة مثل نماذج المحوّلات في تصنيف الغرف والتقسيم الدلالي، إضافة إلى نموذج تحسين الأقنعة، ساهم في رفع جودة النتائج النهائية مقارنة باستخدام نموذج واحد فقط. كما أظهر النظام قدرة جيدة على التعامل مع صور واقعية مختلفة، مما يعكس قابلية التطبيق العملي للفكرة المقترحة.

## تحقيق أهداف البحث:

تم تحقيق الأهداف الأساسية لهذا البحث بدرجة مرضية، حيث نجح النظام في:

* التعرّف على نوع الغرفة تلقائيًا واستخدام هذه المعلومة لدعم مراحل المعالجة اللاحقة.
* استخراج أقنعة دلالية للعناصر المعمارية الأساسية مثل الجدران والأرضيات.
* تحسين دقة هذه الأقنعة باستخدام نموذج متخصص، مما ساهم في تقليل التداخل بين العناصر.
* تطبيق الخامات أو الألوان على العناصر المستهدفة بشكل موضعي، مع الحفاظ على الإضاءة والظلال الأصلية للمشهد.

تعكس هذه النتائج نجاح النظام في تحقيق الهدف الرئيسي المتمثل في دعم تعديل المشاهد الداخلية بطريقة ذكية وتلقائية نسبيًا، مع تقليل الحاجة إلى تدخل يدوي مباشر.

## حدود النظام الحالي:

على الرغم من النتائج الإيجابية، إلا أن النظام الحالي يعاني من بعض القيود. من أبرز هذه القيود الاعتماد على الصور ثنائية الأبعاد (2D) في تمثيل المشهد الداخلي، حيث أظهرت التجارب أن تطبيق الخامات على الصور ثنائية الأبعاد فقط قد يؤدي في بعض الحالات إلى نقص في الواقعية أو ظهور تشوهات بصرية، خاصة عند التعامل مع المنظور، العمق، أو التداخل المعقد بين العناصر.

كما يقتصر النظام الحالي على التعامل مع عناصر معمارية محددة، مثل الجدران والأرضيات، دون الاستفادة الكاملة من جميع الأقنعة الدلالية المتاحة (مثل أقنعة السقف)، إضافة إلى عدم تضمين تقدير كمي دقيق لمساحات العناصر أو تكاليف المواد المستخدمة.

## الآفاق المستقبلية وتطوير العمل:

يفتح هذا المشروع المجال أمام مجموعة واسعة من التوسعات المستقبلية التي يمكن أن تعزز من واقعية النظام وقيمته التطبيقية. ومن أهم هذه التوجهات إدخال نماذج ثلاثية الأبعاد (3D) لتمثيل المشاهد الداخلية، بدل الاعتماد الكامل على الصور ثنائية الأبعاد. إذ إن دمج النماذج ثلاثية الأبعاد أو معلومات العمق يمكن أن يساهم في تمثيل أدق لأبعاد الجدران والأرضيات، ومعالجة مشكلات المنظور، وبالتالي تحسين واقعية تطبيق الخامات والمواد وتقليل مظاهر التزييف البصري.

كما يمكن في الأعمال المستقبلية دمج نظام لتقدير المساحات والكميات، بحيث يتم حساب المساحة الفعلية للجدران والأرضيات بناءً على النموذج ثلاثي الأبعاد أو معلومات العمق، ومن ثم ربط هذه المساحات بالخامات المختارة لتقدير الكميات المطلوبة والتكلفة التقريبية. من شأن هذا التوجه تحويل النظام من أداة عرض بصري فقط إلى أداة دعم قرار تساعد المستخدم في التخطيط والتنفيذ.

إضافة إلى ذلك، يمكن توسيع النظام ليشمل إدارة الأثاث داخل المشهد الداخلي، سواء عبر إضافة عناصر أثاث جديدة، أو تغيير الأثاث الموجود، أو إزالته، مما يسمح بمحاكاة سيناريوهات تصميم داخلي متكاملة تشمل العناصر المعمارية والمفروشات معًا. ويُعد هذا التوسّع خطوة طبيعية نحو بناء نظام تصميم داخلي ذكي شامل.

كما يمكن استثمار الأقنعة الدلالية المتوفرة لعناصر إضافية، مثل السقف، في مراحل لاحقة، وإتاحة خيارات تعديل أوسع للمستخدم بحسب طبيعة العنصر المعماري، كالتفريق بين تغيير اللون أو الخامة للجدران، وتطبيق الخامات فقط للأرضيات، بما يعكس الاستخدام الواقعي لهذه العناصر.

## الخلاصة النهائية:

في الختام، يقدّم هذا البحث إطارًا عمليًا ومرنًا لتعديل المشاهد الداخلية باستخدام الذكاء الاصطناعي، معتمدًا على دمج عدة نماذج متخصصة ضمن خط معالجة واحد. وعلى الرغم من بعض القيود المرتبطة بالتمثيل ثنائي الأبعاد، إلا أن النتائج التي تم تحقيقها تؤكد إمكانية تطوير النظام ليصبح أكثر واقعية وشمولية في المستقبل.

تشير الآفاق المستقبلية المقترحة إلى أن هذا العمل يمكن أن يشكّل أساسًا لأنظمة تصميم داخلي ذكية متقدمة، تجمع بين التعديل البصري، التقدير الكمي، وحساب التكاليف، وتوفر أدوات فعّالة لدعم المستخدمين في اتخاذ قرارات تصميمية دقيقة.

# **الفصل الثامن :المراجع References**

[1] Zhou, Bolei, Aditya Lapedriza, Aude Oliva, Alexei A. Efros, and Antonio Torralba.  
**“Places: A 10 Million Image Database for Scene Recognition.”**  
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017.  
 <https://places2.csail.mit.edu/>

[2] Silberman, Nathan, Derek Hoiem, Pushmeet Kohli, and Rob Fergus.  
**“Indoor Segmentation and Support Inference from RGB-D Images.”**  
European Conference on Computer Vision (ECCV), 2012.  
https://cs.nyu.edu/~silberman/datasets/nyu\_depth\_v2.html

[3] Xie, Enze, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo.  
**“SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers.”**  
Advances in Neural Information Processing Systems (NeurIPS), 2021.  
<https://github.com/NVlabs/SegFormer>

[4] Kirillov, Alexander, Eric Mintun, Nikhila Ravi, et al.  
**“Segment Anything.”**  
arXiv preprint, 2023.  
<https://segment-anything.com>  
<https://github.com/facebookresearch/segment-anything>

[5] Liu, Ze, Yutong Lin, Yue Cao, et al.  
**“Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.”**  
IEEE International Conference on Computer Vision (ICCV), 2021.  
<https://github.com/microsoft/Swin-Transformer>

[6] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.  
**“Deep Residual Learning for Image Recognition.”**  
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.  
<https://arxiv.org/abs/1512.03385>

[7] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, et al.  
**“An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale.”**  
International Conference on Learning Representations (ICLR), 2021.  
<https://arxiv.org/abs/2010.11929>

[8] Kingma, Diederik P., and Jimmy Ba.  
**“Adam: A Method for Stochastic Optimization.”**  
International Conference on Learning Representations (ICLR), 2015.  
<https://arxiv.org/abs/1412.6980>

[9] Zhang, Hongyi, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz.  
**“mixup: Beyond Empirical Risk Minimization.”**  
International Conference on Learning Representations (ICLR), 2018.  
<https://arxiv.org/abs/1710.09412>

[10] Everingham, Mark, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman.  
**“The Pascal Visual Object Classes (VOC) Challenge.”**  
International Journal of Computer Vision (IJCV), 2010.  
<http://host.robots.ox.ac.uk/pascal/VOC/>

## المراجع الخاصة بالمكتبات والأدوات البرمجية

[11] Hugging Face.  
**“SegFormer Model Documentation.”**  
<https://huggingface.co/docs/transformers/model_doc/segformer>

[12] Hugging Face.  
**“Swin Transformer Model Documentation.”**  
<https://huggingface.co/docs/transformers/model_doc/swin>

[13] Meta AI.  
**“Segment Anything Model (SAM) – Official Documentation.”**  
<https://github.com/facebookresearch/segment-anything>

[14] OpenCV.  
**“OpenCV: Open Source Computer Vision Library Documentation.”**  
https://docs.opencv.org/

[15] Scikit-learn.  
**“Classification Metrics and Model Evaluation Documentation.”**  
<https://scikit-learn.org/stable/modules/model_evaluation.html>

"The Pascal Visual Object Classes (VOC) Challenge."  
International Journal of Computer Vision, 2010.

[11] Hugging Face.

"SegFormer Model Documentation."

[12] Hugging Face.

"Swin Transformer Model Documentation."

[13] Meta AI.

"Segment Anything Model (SAM) Official Documentation."

[14] OpenCV.

"OpenCV: Open Source Computer Vision Library Documentation."

[15] Scikit-learn.

"Classification Metrics and Model Evaluation Documentation."