الجمهورية العربية السورية

المعهد العالي للعلوم التطبيقية والتكنولوجيا

قسم المعلوماتية

العام الدراسي 2023/2022

بحث أعد لنيل درجة الماجستير في الهندسة المعلوماتية

اختصاص نظم المعطيات الكبيرة

نموذج لتحسين دقة الفيديو اعتماداً على الشبكات العصبونية العُقدية

A model for video super-resolution using complex-valued neural network

إعداد

م. هيلدا زاهر

إشراف

الدكتور آصف جعفر **الدكتور باسم السهوة**

دمشق 2024

الجمهورية العربية السورية

المعهد العالي للعلوم التطبيقية والتكنولوجيا

قسم المعلوماتية

العام الدراسي 2023/2022

تصريح

أنا الموقع أدناه هيلدا علي زاهر معدّ أطروحة الماجستير التي تحمل العنوان:

نموذج لتحسين دقة الفيديو اعتمادا على الشبكات العصبونية العقدية

أصرح بأن:

– الأعمال والنتائج المعروضة في هذه الأطروحة هي نتيجة جهودي الشخصية وبتوجيه من المشرفين، وأنّ ما عدا ذلك من معلومات ونتائج قد نسبت إلى مصادرها ومؤلفيها، وأشير إلى ذلك في متن النص وفي قائمة المراجع.

– المعطيات والمعلومات المستخدمة في هذه الأطروحة جرى تحصيلها بطرائق سليمة ومشروعة ونُسبت إلى مصادرها في المواضع الملائمة.

– كلّ مكون من مكونات هذه الأطروحة (مقطع نصّي، صورة، مخطط،..) مقتبس من عمل آخر جرى تمييزه بوضوح ونسب إلى مصدره.

– الأعمال والنتائج المعروضة في هذه الأطروحة لم تستخدم سابقًا وليست قيد الاستخدام للحصول على أي شهادة أكاديمية أخرى.

التوقيع

دمشق / /2024

المعهد العالي للعلوم التطبيقية والتكنولوجيا
Higher Institute for Applied Science and Technology

المعهد العالي للعلوم التطبيقية والتكنولوجيا مؤسسة حكومية للتعليم العالي أحدثت بموجب المرسوم التشريعي رقم / 24 / لعام 1983، وذلك بهدف إعداد كوادر علمية متميزة من مهندسين وباحثين للإسهام الفاعل في عملية التطوير العلمي والتنمية في الجمهورية العربية السورية.

يمنح المعهد العالي درجة الإجازة في الهندسة في الاتصالات والمعلوماتية والنظم الالكترونية والميكاترونيكس وعلوم وهندسة المواد وهندسة الطيران. يقبل المعهد العالي لدراسة هذه الاختصاصات شريحة منتقاة من المتفوقين في الشهادة الثانوية من الفرع العلمي. يتيح المعهد العالي أيضاً برامج ماجستير أكاديمي في نظم الاتصالات وفي التحكم والروبوتيك وفي نظم المعطيات الكبيرة ونظم المعلومات ودعم القرار وفي علوم وهندسة المواد وعلوم وهندسة البصريات .وأخيراً، يمنح المعهد العالي درجة الدكتوراه في الاتصالات والمعلوماتية ونظم التحكم والفيزياء التطبيقية .تُحدث في المعهد العالي اختصاصات جديدة بحسب متطلبات سوق العمل وتوجهات البحث والتطوير المحلية والعالمية.

إلى جانب النشاط التعليمي، يمارس المعهد العالي عبر جهود أطره وفعالياته العلمية المختلفة نشاطاً حثيثاً في البحث والتطوير، إذ ينفذ مشاريع ودراسات واستشارات متنوعة لصالح الجهات العامة والخاصة في القطر، كما يتعاون مع جهات خارج القطر في بعض المشاريع البحثية والتطويرية. يسعى المعهد أيضاً، عبر دورات تدريبية نظرية وعملية متاحة للقطاعين العام والخاص وللأفراد، إلى إفادة أوسع فئة من المهتمين من إمكانيات أطره العلمية ومختبراته. واستكمالاً لدوره الرائد في مجال التعليم ونشر العلم، ينشر المعهد العالي كتباً علمية عالية المستوى من نتاج أطره، منها ما هو تدريسي يوافق المناهج في المعهد العالي ويفيد شريحة واسعة من الطلاب الجامعيين عموماً، ومنها ما هو علمي ثقافي. يتيح المعهد العالي بعضاً من منشوراته على موقعه على شبكة الانترنيت، كما يتيح امكانية الاطلاع على رسائل الماجستير والدكتوراه المنفذة في المعهد العالي وعلى بعض منشورات طلبه وأطره من المقالات العلمية.

المعهد العالي للعلوم التطبيقية والتكنولوجيا، الجمهورية العربية السورية، دمشق، ص.ب31983
Higher Institute for Applied Sciences and Technology – HIAST
P. O. Box 31983, Damascus, Syrian Arab Republic
هاتف – 00963115123819 فاكس00963115140761
بريد الكتروني contact@hiast.edu.sy
موقع الكتروني www.hiast.edu.sy

كلمة شكر

إلى من كان لدعمه وعطائه الدور الأبرز في إتمام هذا العمل، أتوجه بخالص الشكر وجزيل الامتنان إلى الدكتور آصف جعفر والدكتور باسم السهوة، على إشرافهما الكريم وإخلاصهما في تقديم النصح والتوجيه. إن وقتهما الثمين وإرشاداتهما الهادفة ومتابعتهما المستمرة كانت خير معين لتحقيق هذا الإنجاز، وأساساً متيناً لنجاح هذا البحث.

كما أعرب عن بالغ شكري وتقديري للدكتور عمر حمدون، الذي لم يبخل عليّ بمساعدته وتوجيهاته القيّمة. كان دعمه مصدر إلهام ودافعاً لتجاوز العديد من التحديات، وقد أسهمت توجيهاته السديدة في إثراء مسيرة البحث بشكل لا يُنسى.

ولا يفوتني أن أقدم عظيم الشكر والامتنان إلى أساتذتي الأفاضل في المعهد العالي للعلوم التطبيقية والتكنولوجيا. لقد كان لدعمهم وتوجيههم الأثر العميق في مسيرتي الأكاديمية، وكانوا نموذجاً يُحتذى في العلم والأخلاق، مما سيبقى دائماً موضع فخر واعتزاز في كل خطوة أخطوها مستقبلاً.

إهداء

إلى من سكنوا القلب وكانوا النور الذي أنار لي الدرب، إلى عائلتي الحبيبة التي كانت معي في كل خطوة وفي كل لحظة تعب وسهر، أُهدي ثمرة هذا الجهد. إلى أعزّ الناس على قلبي، **محمود وقمر**، اللذين كانا دوماً الدافع الأكبر للاستمرار، وملاذي الآمن حين تثاقلت عليّ الصعاب، أُهديكما هذا العمل بكل ما فيه من أمل وتعب.

إلى إخوتي الأعزاء، الذين كانوا سنداً لي في أوقات الشدة ومصدر عزمي في أوقات الضعف، وإلى ذكريات والديّ الغاليين، اللذين رحلا عن الدنيا وخلّفا في روحي قوة وإرثاً لا يُنسى من المحبة والأمل، أسأل الله أن يتغمدهما برحمته، وأن يجعل هذا العمل صدقةً جاريةً في ميزان حسناتهما.

لكم جميعاً، أهدي هذا العمل، ممتنةً ومحبةً إلى ما لا نهاية.

الخلاصة

تُعد تقنية الدقة الفائقة (Super-Resolution) من التقنيات المهمة في معالجة الصور والفيديوهات، حيث تهدف إلى تحسين دقة الصور منخفضة الجودة لتصبح صورًا عالية الدقة، مما يسمح بتعزيز التفاصيل والوضوح. تستخدم معظم النماذج لتحسين دقة الصور والفيديوهات ,ولاسيما تلك المبنية باستخدام الشبكات العصبونية ,القيم الحقيقية في تمثيل صورة الدخل وما يليها من معالجات لاحقة، مما يجعلها تتجاهل معلومات الطور التي قد تكون مهمة في تمثيل الصورة بدقة أعلى.

ولذلك، يهدف هذا البحث إلى استكشاف إمكانيات الشبكات العصبونية العقدية، حيث يُمكّن هذا التعديل من الاستفادة من معلومات الطور بالإضافة إلى السعة، مما يعزز من قدرة النموذج على تمثيل التفاصيل الدقيقة بشكل أفضل .قمنا بتعديل خوارزميات مبنية باستخدام الشبكات العصبونة ومعتمدة على قيم حقيقية مثل SRCNN وFRVSR وTecoGAN لتصبح شبكات تعتمد على القيم العُقدية، وقمنا بتقييم أدائها باستخدام مجموعات بيانات مرجعية متنوعة تشمل مشاهد طبيعية وحضرية، بهدف قياس مدى تحسين جودة الصور والفيديوات المولدة. اعتمد التقييم على مقاييس شائعة مثل PSNR و SSIMلتقييم الدقة، بالإضافة إلى مقاييس زمنية مثل tof وtLp100 لضمان التناسق الزمني عبر الإطارات المتتابعة .أظهرت النتائج أن الشبكات العصبونية ذات القيم العُقدية حققت تحسنًا ملحوظًا مقارنة بالنماذج ذات القيم الحقيقية، حيث قدمت أداءً أفضل في تحسين دقة الصور والتناسق الزمني للفيديوات دون زيادة كبيرة في عدد البارامترات. يدل ذلك على أن الاستفادة من معلومات الطور، عبر استخدام القيم العُقدية، يمكن أن يُحسّن جودة النتائج ويعزز من فعالية خوارزميات الدقة الفائقة في معالجة الصور والفيديوات .

الكلمات المفتاحية: الدقة الفائقة، الشبكات العصبونية العميقة، القيم العُقدية، معلومات الطور، SRCNN، FRVSR، TecoGAN، PSNR، SSIM.

Abstract

Super-Resolution (SR) is a significant technique in image and video processing, aiming to enhance the resolution of low-quality images, transforming them into high-resolution ones, thereby improving detail and clarity. Most models for image and video resolution enhancement, especially those based on neural networks, rely on real values for representing input images and subsequent processing stages. This approach often overlooks phase information, which could be essential for achieving more accurate image representation.

This research aims to explore the potential of complex-valued neural networks, enabling the incorporation of phase information alongside magnitude, thus enhancing the model's ability to represent fine details more effectively. We modified existing algorithms based on real-valued neural networks, such as SRCNN, FRVSR, and TecoGAN, to operate using complex values. The performance of these models was evaluated using diverse benchmark datasets encompassing natural and urban scenes, aiming to assess the quality improvement of the generated images and videos. The evaluation employed commonly used metrics, such as PSNR and SSIM, for accuracy, as well as temporal metrics like tof and tLp100 to ensure temporal consistency across consecutive frames.

The results demonstrated that complex-valued neural networks achieved significant improvements compared to their real-valued counterparts, delivering superior performance in image resolution enhancement and temporal consistency in videos without incurring substantial increases in parameter count. This indicates that leveraging phase information through complex values can enhance the quality of results and improve the efficiency of super-resolution algorithms in image and video processing.

Keywords: Super-Resolution, Deep Neural Networks, Complex Values, Phase Information, SRCNN, FRVSR, TecoGAN, PSNR, SSIM.

المحتويات

الفصل الأول – الإطار العام للبحث 1
1.1 – خلفية عامة 1
2.1 – مشكلة البحث 2
3.1 – أهداف البحث 2
4.1 – أسئلة البحث 2
5.1 – أثر البحث وأهميته 2
6.1 – هيكل الأطروحة 3
الفصل الثاني – مراجعة الأدبيات 4
1.2 – تقنيات تحسين دقة الفيديو التقليدية 4
1.1.2 – الأساليب القائمة على الاستيفاء 4
2.1.2 – الأساليب القائمة على البناء 4
3.1.2 – الأساليب القائمة على التعلم **5**
4.1.2– الخوارزميات القائمة على التعلم العميق 5
1.4.1.2 – الطرق المعتمدة على الشبكات العصبونية التلافيفية (CNNs) 6
2.4.1.2 – الطرق المعتمدة على الشبكات التكرارية (RNNs) 7
3.4.1.2 – الطرق المعتمدة على شبكات الخصوم التوليدية (GANs) 8
4.4.1.2 – الطرق المعتمدة على نماذج قائمة على المحولات (Transformers) 9
2.2 – الشبكات العصبونية ذات القيم العُقدية (CVNNs) 10
1.2.2 – البنية الرياضية للشبكات العصبونية ذات القيم العُقدية 10
2.2.2 – العمليات الرياضية الأساسية في CVNNs 11
3.2.2 – أهمية CVNNs في تحسين الدقة الفائقة للفيديو 11
4.2.2 أهمية الطور في تحسين دقة الفيديو 11
3.2 –مقاييس التقييم ومجموعات البيانات 12
1.3.2 – مقاييس التقييم 12
1.1.3.2 – نسبة الإشارة إلى الضوضاء (PSNR) 12
2.1.3.2 –مؤشر التشابه الهيكلي (SSIM) 12

3.1.3.2 تشابه الرقعة الصورية الإدراكية المتعلمة (LPIPS) 13
2.3.2 مجموعات البيانات المرجعية 14
1.2.3.2 - Vid4 14
2.2.3.2 - REDS 14
3.2.3.2 - Vimeo-90K 14
الفصل الثالث – النموذج المقترح 15
1.3 - SRCNN 15
1.1.3 – بنية خوارزمية SRCNN 16
2.1.3 – تدريب الشبكة 18
2.3 - FRVSR 19
1.2.3 البنية العامة لخوارزمية FRVSR 20
1.2.3 الآلية التفصيلية لعمل FRVSR 20
1.2.3 – تدريب الشبكة ودوال الخسارة 24
3.3 - TecoGAN 25
1.3.3 كيفية عمل شبكات GAN التوليدية التنافسية 25
2.3.3 البنية العامة لخوارزمية TecoGAN 26
1.2.3.3 – الشبكة التوليدية (SRNet) 26
2.2.3.3 المُميز المكاني–الزماني 27
3.3.3 توابع الخسارة المستخدمة في تدريب TecoGAN 28
4.3 – بنية شبكة FNet والتحويل إلى القيم العُقدية 30
1.4.3 بنية شبكة FNet 31
2.4.3 تحويل شبكة FNet إلى شبكة عصبية ذات قيم عُقدية (CVNNs) 31
1.2.4.3 آلية التحويل 32
الفصل الرابع – التطبيق العملي و النتائج 34
1.4 نتائج خوارزمية SRCNN مع القيم العُقدية 34
1.1.4 – إعداد التجربة 34
2.1.4 – مجموعة التدريب ومقاييس التقييم 34
3.1.4 – التدريب 35
4.1.4 تحليل النتائج 36
5.1.4 – مقارنة النتائج البصرية 37

2.4 إعداد التجربة لخوارزميات FRVSR وTecoGAN 37
1.2.4 - مجموعة التدريب 37
2.2.4 إعدادات التدريب 38
3.2.4 - المقاييس المستخدمة لتقييم FRVSR وTecoGAN 40
3.4 - نتائج FRVSR وComplex-FRVSR 40
4.4 - نتائج TecoGAN وComplex-TecoGAN 41
5.4 - التحليل العام 46
الفصل الخامس – الخاتمة والآفاق المستقبلية 47
المراجع 48

قائمة الأشكال

الشكل 1- مقارنة خوارزمية SRCNN مع طريقة الاستيفاء التكعيبي وطريقة الترميز المتناثرSC 16
الشكل 2- المراحل الثلاث لخوارزميةSRCNN 18
الشكل 3- مخطط تدفق عملية FRVSR موضحًا شبكتي FNet وSRNet ، بالإضافة إلى آليات الحركة والخسارة المستخدمة أثناء التدريب..... 20
الشكل 4- التحويل من الفضاء إلى العمق 22
الشكل 5- المخطط العام لبنيةTecoGAN 26
الشكل 6- المدخلات المستخدمة في المُميز المكاني-الزماني فيTecoGAN 27
الشكل 7- آلية عمل تابع خسارةPing-Pong 29
الشكل 8- بنية شبكة التدفق البصريFNet 31
الشكل 9- المقارنة البصرية بين SRCNN وCSRCNN 37
الشكل 10- منحنى تدريب خوارزميةFRVSR 38
الشكل 11- منحنى تدريبComplex-FRVSR 39
الشكل 12- منحنى تدريب خوارزميةTecoGAN 39
الشكل 13- منحنى تدريبComplex-TecoGAN 40
الشكل 14- المقارنة البصرية بين FRVSR وComplex-FRVSR 43
الشكل 15- المقارنة البصرية بين TecoGAN وComplex-TecoGAN 46

قائمة الجداول

الجدول 1- جدول مقارنة PSNR لخوارزمية SRCNN والنسخة المعدلةC-SRCNN 35
الجدول 2- جدول مقارنة SSIM لخوارزمية SRCNN والنسخة المعدلةC-SRCNN 36
الجدول 3- جدول PSNR مع تغيير المرشحات وعدد البارامترات لكل شبكة 37
الجدول 4- نتائج خوارزمية FRVSR والنسخة المعدلة Complex-FRVSR على مجموعة الاختبار vid4 41
الجدول 5- نتائج خوارزمية FRVSR والنسخة المعدلة Complex-FRVSR على مجموعة الاختبار tos3..... 42
الجدول 6- نتائج خوارزمية TecoGAN والنسخة المعدلةComplex-TecoGAN على مجموعة الاختبار vid4 44
الجدول 7- نتائج خوارزمية TecoGAN والنسخة المعدلة Complex-TecoGAN على مجموعة الاختبار tos3.45

الاختصارات والمصطلحات

الاختصار	المصطلح باللغة العربية	المصطلح باللغة الإنجليزية
CNNs	الشبكات العصبونية التلافيفية	Convolutional Neural Networks
RNNs	الشبكات العصبونية التكرارية	Recurrent Neural Networks
GANs	شبكات الخصوم التوليدية	Generative Adversarial Networks
CVNNs	الشبكات العصبونية ذات القيم العُقدية	Complex-Valued Neural Networks
PSNR	نسبة الإشارة إلى الضوضاء القصوى	Peak Signal-to-Noise Ratio
SSIM	مؤشر التشابه الهيكلي	Structural Similarity Index Measure
LPIPS	تشابه الرقعة الصورية الإدراكية المتعلمة	Learned Perceptual Image Patch Similarity
TecoGAN	شبكة الخصوم التوليدية ذات التناسق الزمني	Temporally Coherent Generative Adversarial Network
SRCNN	الشبكة العصبونية التلافيفية لتحسين الدقة الفائقة	Super Resolution Convolutional Neural Network
VSRNet	شبكة تحسين دقة الفيديو باستخدام الشبكات التلافيفية	Video Super-Resolution Network
VESPCN	شبكة تكبير الفيديو وتحسينه	Video Enhancement and Super-Resolution Network
EDVR	الاستعادة الفيديوية المشوهة المحسنة	Enhanced Deformable Video Restoration
RBPN	شبكة الإسقاط الخلفي المتكررة	Recurrent Back-Projection Network
STTN	شبكة المحولات الزمنية المكانية	Spatio-Temporal Transformer Network
REDS	مجموعة بيانات مشاهد واقعية ومتنوعة	The realistic and dynamic scenes
Vimeo-90K	مجموعة بيانات فيديوهات من منصة فيميو	Vimeo-90K

Set5	مجموعة بيانات تحتوي على 5صور لاختبارات الدقة	Set5
Set14	مجموعة بيانات تحتوي على 14صورة لاختبارات الدقة	Set14
BSD100	مجموعة بيانات تحتوي على 100صورة لتحسين الدقة	BSD100
Urban100	مجموعة بيانات تحتوي على 100صورة حضرية	Urban100
MSE	الخطأ المتوسط التربيعي	Mean Squared Error
FNet	شبكة تقدير التدفق البصري	Flow Estimation Network
SRNet	شبكة تحسين الدقة	Super-Resolution Network
T91	مجموعة تدريب تتألف من 91صورة	T91
PSNR	نسبة الذروة للإشارة إلى الضوضاء	Peak Signal-to-Noise Ratio
SSIM	مؤشر التشابه الهيكلي	Structural Similarity Index
LPIPS	تشابه التصور التعلمي لرقع الصورة	Learned Perceptual Image Patch Similarity
TOF	التدفق البصري الزمني	Temporal Optical Flow
TLP100	مؤشر LPIPS الزمني لـ 100 إطار	Temporal LPIPS for 100 frames

الفصل الأول

الإطار العام للبحث

نبيّن في هذا الفصل مجال البحث ومشكلة البحث المطلوب العمل عليها والمقاربات التي اعتمدت في معالجة هذه المشكلة ونحدد المقاربة التي ستعتمد في البحث.

1.1– خلفية عامة

إن عملية تصوير مشهد ما باستخدام الكميرا هي عملية تقطيع للإشارة المستمرة لهذا المشهد وذلك عند مستوى الحساس لهذه الكميرا. تؤدي عملية التقطيع هذه إلى ضياع الترددات العالية (الحواف والتفاصيل الدقيقة) ضمن هذه الإشارة المستمرة والتي تظهر على شكل تردي في جودة أو دقة الصورة أو الفيديو أو نوع من الضبابية. لمعالجة هذه المشكلة ظهرت تقنيات تحسين دقة الفيديو (Video Super-resolution) والتي تهدف لاسترجاع الترددات العالية المفقودة أثناء عملية التصوير.

يستخدم مصطلح الدقة الفائقة أو تحسين الدقة (Super-Resolution) لوصف إجرائية الحصول على صورة أو مجموعة إطارات/صور (فيديو) عالية الدقة (High-Resolution) انطلاقاً من مجموعة من الصور منخفضة الدقة (Low-Resolution). يعتبر تحسين دقة الفيديومن المجالات الحيوية في تكنولوجيا الوسائط المتعددة والاتصالات والذي زادت أهميته خصوصاً مع التطور السريع في تقنيات العرض، بما في ذلك شاشات التلفاز عالية الدقة وأجهزة العرض الرقمية. بالإضافة إلى ذلك ، تُستخدم الدقة العالية في المجال الطبي لتحليل البيانات المرئية بدقة، والذي بدوره يؤدي إلى زيادة دقة التشخيص الطبي . أما في مجال المراقبة الأمنية، فإن تحسين دقة الفيديو يسهم في زيادة وضوح الصور الملتقطة، مما يساعد على التعرف الدقيق على العناصر في الفيديوهات.

في السنوات الأخيرة، كانت معظم الأبحاث تركز على استخدام الشبكات العصبونية ذات القيم الحقيقية لتحسين دقة الفيديو، وبالتالي فإن هذه الأساليب تركز بشكل كبير على تحليل السعة دون الاستفادة الكاملة من معلومات الطور. تمثل السعة مقدار تواتر تردد مكاني ما ضمن الصورة أما الطور فهو يمثل مواقع هذه الترددات المكانية بدقة ، وبالتالي يحتوي على معلومات حاسمة يمكن أن تسهم في تحسين التفاصيل الدقيقة للفيديو. من هنا تنبع أهمية البحث في استخدام الشبكات العصبونية ذات القيم العُقدية، التي تمكن من استغلال معلومات الطور والسعة معًا، مما يفتح آفاقًا جديدة لتحسين دقة الفيديو.

2.1– مشكلة البحث

إن على الرغم من التقدم الكبير في تحسين دقة الفيديو باستخدام الشبكات العصبونية، لا تزال هناك فجوة معرفية كبيرة فيما يتعلق بكيفية الاستفادة المثلى من جميع المعلومات المتاحة في إشارة الفيديو. تستخدم الشبكات العصبونية التقليدية معلومات السعة فقط لإعادة بناء الصور، مما يؤدي إلى إهمال معلومات الطور التي تعد أساسية لتمثيل المعلومات المكانية بدقة. هذا الإهمال لمعلومات الطور يؤدي إلى قصور في الأداء العام للنماذج ويحد من قدرتها على تحسين دقة الفيديو بشكل كامل.

لذلك، تكمن مشكلة البحث في كيفية إدماج معلومات الطور بشكل فعال في نماذج الشبكات العصبونية لتحسين دقة الفيديو، مما يعزز من جودة الصورة الناتجة ويزيد من دقتها.

3.1– أهداف البحث

يهدف هذه البحث إلى تحقيق ما يلي:

1. تطوير نموذج جديد لتحسين دقة الفيديو يعتمد على الشبكات العصبونية ذات القيم العُقدية، والذي يمكنه الاستفادة من معلومات الطور والسعة بشكل متكامل.
2. إجراء مقارنة بين أداء النموذج الجديد ومقارنته بالنماذج التقليدية التي تعتمد على القيم الحقيقية فقط.
3. دراسة تأثير دمج معلومات الطور مع معلومات السعة على أداء نموذج تحسين جودة الفيديو من حيث زيادة الدقة والكلفة الحسابية.

4.1– أسئلة البحث

يسعى هذا البحث للإجابة على الأسئلة البحثية التالية:

- كيف يمكن استخدام معلومات الطور في تحسين دقة الفيديو؟
- ما هي الفروقات في الأداء بين الشبكات العصبونية ذات القيم الحقيقية والشبكات العصبونية ذات القيم العُقدية؟
- ما هو تأثير دمج معلومات الطور مع معلومات السعة على جودة الفيديو الناتج من حيث الدقة والتفاصيل المكانية؟
- ما هي الفروقات في التعقيد الحسابي بين الشبكات العصبونية ذات القيم الحقيقية والشبكات العصبونية ذات القيم العُقدية؟

5.1– أثر البحث وأهميته

تبرز أهمية هذا البحث من خلال تقديمه لإسهامات نظرية وعملية. من الناحية النظرية، يساهم البحث في تعميق فهم الشبكات العصبونية ذات القيم العُقدية ودورها في معالجة معلومات الطور في الإشارات الرقمية. ستقدم نتائج هذا البحث رؤى جديدة حول كيفية تحسين الدقة العالية للفيديو باستخدام تقنيات متقدمة تستفيد من المعلومات المتاحة بشكل كامل.

أما من الناحية العملية، فإن هذا البحث قد يساهم في تطوير تقنيات جديدة تُستخدم في مجموعة واسعة من التطبيقات التي تتطلب فيديوهات عالية الدقة. سواء في مجالات البث التلفزيوني أو الطب أو المراقبة. علاوة على ذلك، قد يسهم هذا البحث في إيجاد حلول جديدة للتحديات المتعلقة بتحسين الدقة العالية للفيديو والتي لم تحل بشكل كامل بعد.

6.1– هيكل الأطروحة

تتألف هذه الأطروحة من خمسة فصول أساسية تهدف إلى تغطية موضوع تحسين دقة الفيديو باستخدام خوارزميات الشبكات العصبونية التقليدية والعُقدية، بدءًا من تقديم الإطار العام وصولاً إلى عرض النتائج والخاتمة.

–الفصل الأول – الإطار العام للبحث

يهدف هذا الفصل إلى تقديم خلفية عامة عن موضوع البحث، وتحديد مشكلة البحث بشكل واضح، إضافة إلى توضيح أهداف البحث وأسئلته. كما يتطرق إلى أهمية البحث وأثره، وينتهي بتقديم موجز عن هيكل الأطروحة وترتيب محتوياتها.

–الفصل الثاني – مراجعة الأدبيات

يستعرض هذا الفصل الأدبيات السابقة المتعلقة بتقنيات تحسين دقة الفيديو، ويغطي ثلاثة جوانب رئيسية: الأساليب التقليدية مثل الاستيفاء وإعادة البناء، ثم الانتقال إلى الأساليب القائمة على التعلم والتعلم العميق. كما يناقش هذا الفصل الشبكات العصبونية ذات القيم العُقدية (CVNNs)، وأهميتها في تحسين الدقة الفائقة للفيديو. يُختتم الفصل بعرض مقاييس التقييم ومجموعات البيانات المرجعية المستخدمة في تحسين دقة الفيديو.

–الفصل الثالث – النموذج المقترح

يركز هذا الفصل على عرض الخوارزميات المقترحة لتحسين دقة الفيديو، حيث يتم عرض خوارزمية SRCNN كنموذج تقليدي، يليها خوارزميات FRVSR وTecoGAN التي تعتمد على التكرارية (التسلسل الزمني للإطارات) والشبكات التوليدية التنافسية حيث يتم بشكل أساسي شرح بنية شبكة FNet المستخدمة ضمن هاتين الخوارزميتين لحساب الإنزياحات بين الأطر المتتالية، وطرح آلية تحويلها إلى شبكة ذات قيم عُقدية بهدف تحسين دقة النتائج والحفاظ على التناسق الزمني والمكاني في الفيديو.

–الفصل الرابع – التطبيق العملي والنتائج

يعرض هذا الفصل خطوات التطبيق العملي على خوارزميات SRCNN، FRVSR وTecoGAN، بدءًا من إعداد التجارب وصولاً إلى تحليل النتائج. يتم التطرق إلى إعدادات التدريب والمقاييس المستخدمة في تقييم الأداء، مع التركيز على مقارنة الأداء بين النسخ الأصلية وتلك المعتمدة على القيم العُقدية، مما يوضح التحسينات المحققة في جودة الفيديو وتناسق الإطارات.

–الفصل الخامس – الخاتمة والآفاق المستقبلية

يُختتم البحث بتقديم ملخص عام للنتائج المستخلصة من الدراسة، مع توضيح الإسهامات النظرية والعملية لهذا البحث. كما يتم عرض بعض الاقتراحات والتوصيات لمواصلة البحث في هذا المجال وتطوير تطبيقات مستقبلية لهذه الخوارزميات في تحسين دقة الفيديو.

يتبع الأطروحة قسم المراجع، الذي يضم قائمة بالمصادر الأكاديمية والكتب والمقالات التي تم الاستناد إليها خلال إعداد البحث.

الفصل الثاني

مراجعة الأدبياتLiterature review

يلخص ويقارن وينتقد هذا الفصل أهم المصادر والمقالات العلمية الأكاديمية المتصلة بمشكلة البحث والمسائل المنبثقة عنها

سنستعرض في هذا الفصل مراجعة شاملة للخوارزميات الاساسية المستخدمة في تحسين دقة الفيديو ، مع التركيز على التطورات الحديثة في التعلم العميق، مثل الشبكات العصبونية التلافيفية (CNNs)، الشبكات التكرارية (RNNs)، وشبكات الخصومة التوليدية (GANs). بالإضافة إلى ذلك،سيتم مناقشة الشبكات العصبونية ذات القيم العُقدية (CVNNs) التي تمثل تطورًا جديدًا في هذا المجال.

1.2– تقنيات تحسين دقة الفيديو التقليدية

يمكن تقسيم تقنيات تحسين دقة الفيديو إلى ثلاث فئات:

1.1.2- الأساليب القائمة على الاستيفاء

تعتبر الطرق القائمة على الاستيفاء من أبسط الأساليب المستخدمة لتحسين دقة الفيديو. تعتمد هذه الطرق على خوارزميات مباشرة مثل الاستيفاء بأقرب الجيران، والاستيفاء الثنائي الخطي، والاستيفاء الثنائي التكعيبي لتكبير إطارات الفيديو. ومع ذلك، فإن هذه الطرق غالبًا ما تؤدي إلى صور ضبابية وتفشل في استعادة التفاصيل الدقيقة[1,2] يعود ذلك إلى أن هذه الأساليب لا تأخذ في الحسبان الخصائص الأساسية للصورة، مثل الحواف والتغيرات في الألوان والقوام (texture)، والتي تشكل ما يُعرف بـ "هيكل الصورة". هذه الخصائص هي ما تعطي الصورة تفاصيلها ودقتها. باستخدام الاستيفاء فقط، لا يمكن لهذه الأساليب استنتاج أو إعادة بناء هذه المكونات بشكل صحيح، مما يؤدي إلى فقدان جودة الصورة. ورغم بساطتها، لا تزال الطرق القائمة على الاستيفاء معيارًا أساسيًا للمقارنة في مجال أبحاث تحسين دقة الفيديو.

2.1.2– الأساليب القائمة على إعادة البناء

تهدف الأساليب القائمة على إعادة البناء إلى استعادة التفاصيل عالية التردد من خلال الاستفادة من معرفة مسبقة حول إشارة الفيديو. تعتمد هذه الأساليب على تقنيات التنظيم (Regularization) لفرض النعومة أو الحفاظ على الحواف في الفيديو الُعاد بناؤه.

يُعد تنظيم التباين الكلي (Total Variation Regularization) من بين أبرز هذه الأساليب، حيث يعمل على تقليل الضوضاء مع الحفاظ على الحواف الهامة في الصورة [3].
بالإضافة إلى ذلك، تُستخدم الطرق القائمة على الأمثلة، مثل تلك التي اقترحها Freeman وآخرون [4]، حيث تستفيد من قاعدة بيانات تحتوي على أزواج من المقاطع منخفضة وعالية الدقة لتوجيه عملية إعادة البناء. هذه الأساليب قادرة على استعادة التفاصيل بدقة أعلى مقارنةً بأساليب الاستيفاء، إلا أنها تتطلب تكلفة حسابية أعلى.

تشمل أساليب إعادة البناء أيضًا تحسين دقة الفيديو المعتمد على التعويض الحركي (Motion Compensation)، والذي يستفيد من الارتباط الزمني بين إطارات الفيديو عبر استخدام تقنيات مثل تقدير التدفق البصري (Optical Flow) لمحاذاة الإطارات المتتابعة، مما يسمح بتجميع المعلومات عبر عدة إطارا [7]. يُحسّن هذا النهج من جودة إعادة البناء، لكنه غالبًا ما يعاني من عدم الدقة في تقدير الحركة، مما يؤدي إلى ظهور آثار مثل الأشباح (Ghosting) والضجيج.

3.1.2– الأساليب القائمة على التعلم

قدمت الأساليب القائمة على التعلم تقدمًا كبيرًا مقارنةً بأساليب الاستيفاء وإعادة البناء. تعتمد هذه التقنيات على تعلم تحويلٍ من المقاطع منخفضة الدقة إلى المقاطع عالية الدقة باستخدام مجموعة بيانات تدريبية [5,6]. وعلى الرغم من أنها تحقق أداءً محسّنًا، فإنها عادةً ما تكون مكلفة من الناحية الحسابية[8,9] . كما أن فعاليتها قد تكون محدودة عندما تواجه بيانات تختلف عن تلك الموجودة في مجموعة التدريب، مما يقلل من قابليتها للتطبيق في الحالات الواقعية المتنوعة.

ومع التقدم السريع في مجال التعلم العميق، ظهرت خوارزميات جديدة ذات أداء متفوق في تحسين دقة الفيديو، حيث أصبحت النماذج قادرة على تعلم تحويلات معقدة من تسلسلات الفيديو منخفضة الدقة إلى عالية الدقة[21,22] .

4.1.2– الخوارزميات القائمة على التعلم العميق

أحدث صعود التعلم العميق نقلة نوعية في تحسين دقة الفيديو، إذ أصبحت الخوارزميات الحديثة تعتمد على القدرة الهائلة للشبكات العصبونية في تعلم الأنماط المعقدة واستخراج التفاصيل الدقيقة من البيانات. ومع تطور تقنيات التعلم العميق، ظهرت خوارزميات جديدة تحقق تحسينات ملحوظة في جودة الفيديو. يمكن تقسيم هذه الأساليب بشكل عام إلى:

- **الطرق المعتمدة على الشبكات العصبونية التلافيفية (Convolutional Neural Networks – CNNs)**: مثل شبكة SRCNN [21]، والتي تستخرج الميزات من الصور منخفضة الدقة لإعادة بناء صور عالية الدقة.
- **الطرق المعتمدة على الشبكات العصبونية المتكررة (Recurrent Neural Networks – RNNs)**: تهتم بالتسلسل الزمني للإطارات، مما يساعد في معالجة الفيديو بشكل متكامل مع الحفاظ على استمرارية التفاصيل[11].

- **الطرق المعتمدة على شبكات الخصوم التوليدية (Generative Adversarial Networks – GANs)**: مثل[13] TecoGAN، والتي تستخدم شبكتين متنافستين لتوليد فيديوهات ذات جودة عالية مع المحافظة على الاتساق الزمني بين الإطارات.

- **الطرق المعتمدة على نماذج قائمة على المحولات (Transformers)**: تُستخدم هذه النماذج بشكل متزايد في تحسين دقة الفيديو لما تمتلكه من قدرة على تعلم العلاقات المعقدة بين أجزاء الإطارات المختلفة[19 ,18].

وفيما يلي، سيتم استعراض كل من هذه الأساليب بالتفصيل لتوضيح كيفية عملها ومساهمتها في تحسين دقة الفيديو.

1.4.1.2– الطرق المعتمدة على الشبكات العصبونية التلافيفية (CNNs)

تعتبر الطرق المبنية باستخدام الشبكات العصبونية التلافيفية (CNNs) هي الأساس للعديد من خوارزميات تحسين الدقة الحديثة. تتميز هذه الشبكات بقدرتها على تعلم التمثيلات الهرمية للصور والفيديوهات، مما يسمح لها بتحسين التفاصيل بشكل كبير.

ومن أوائل النماذج التي طُوّرت باستخدام CNN لتحسين الدقة الفائقة للصور هو نموذج SRCNN (شبكة إعادة بناء الصور التلافيفية)، حيث يستخدم لتحسين جودة الصور منخفضة الدقة من خلال تعلم تحويلات مباشرة لإنتاج صور ذات دقة أعلى [21]. ويمثل هذا النموذج خطوةً أساسيةً في تطوير تقنيات CNN لتحسين الدقة.

- **SRCNN** (Super-Resolution Convolutional Neural Network): يُعد نموذج SRCNN أول نموذج مستخدم لتحسين دقة الصور منخفضة الجودة باستخدام الشبكات العصبونية التلافيفية CNN، ويُعتبر أساسًا مهمًا في تطوير خوارزميات زيادة الدقة للصور والفيديوهات [21] .
- **VSRNet** (Video Super-Resolution Network): يعد هذا النموذج من أوائل النماذج المبنية باستخدام شبكات CNN لزيادة دقة الفيديو. يعمل VSRNet على معالجة كل إطار بشكل مستقل باستخدام شبكة تلافيفية قبل دمجه مع الإطارات المجاورة. على الرغم من فعاليته في تحسين دقة الإطارات الفردية، إلا أنه يفتقر إلى نمذجة العلاقات الزمنية بين الإطارات المتتالية، مما يؤدي إلى فقدان التماسك الزمني وإمكانية ظهور تشوهات زمنية عند وجود حركة معقدة [8].
- **VESPCN** (Video Enhancement and Super-Resolution through Spatio-Temporal Convolutional Networks): قدمت VESPCN تحسينات كبيرة على VSRNet من خلال إضافة طبقة تحويل تحت-بكسل (Sub-pixel) زمانية مكانية، مما ساهم في تعزيز التماسك الزمني بين الإطارات وتسريع العملية. ومع ذلك، تبقى مسألة تعويض الحركة بدقة تحديًا كبيرًا. يرجع ذلك إلى تعقيد أنماط الحركة في الفيديو، حيث يمكن أن تكون سريعة، عشوائية، أو متعددة الاتجاهات. إن تقدير الحركة بشكل غير دقيق يمكن أن يؤدي إلى ظهور تشوهات في الإطارات المُعززة، مثل الضبابية أو التموجات، خاصة في المشاهد التي تحتوي على تفاصيل دقيقة أو تغيرات مكانية كبيرة [9].

- **EDVR** (Enhanced Deformable Video Restoration): يقوم EDVR بدمج التفافات مشوهة للتعامل بشكل أفضل مع تعويض الحركة المعقدة. على الرغم من أن هذه الخوارزمية تحسن من دقة التفاصيل المكانية، إلا أن التعامل مع الحركات السريعة أو غير المتوقعة لا يزال يمثل تحديًا [10].

سلبيات الشبكات العصبونية التلافيفية (CNNs):

1. نقص التماسك الزمني (temporal inconsistency): تظهر هذه المشكلة بشكل رئيسي في النماذج التقليدية من CNNs التي تعالج كل إطار في الفيديو بشكل مستقل، دون الاستفادة من المعلومات الزمنية بين الإطارات المتتالية. هذا التجاهل للعلاقات الزمنية يمكن أن يؤدي إلى فقدان التماسك الزمني وظهور تشوهات في الفيديو النهائي. ومع ذلك، فإن بعض النماذج الأحدث، مثل VESPCN، تحاول معالجة هذه المشكلة من خلال دمج المعلومات من الإطارات المجاورة لتحسين التماسك الزمني.
2. الحساسية لتعويض الحركة: تعتمد CNNs بشكل كبير على تقدير الحركة بين الإطارات. إذا كان تقدير الحركة غير دقيق، خاصة في المشاهد ذات الحركات المعقدة، فإن ذلك يمكن أن يؤدي إلى ظهور عيوب وتشوهات في الفيديو المعزز.
3. الاعتماد على التبعية المكانية فقط: تميل CNNs إلى التركيز بشكل رئيسي على العلاقات المكانية داخل الإطار الواحد، مما يجعلها أقل كفاءة في التعامل مع التسلسلات الزمنية المعقدة التي تتطلب نمذجة دقيقة للعلاقات بين الإطارات المتتابعة. وعلى الرغم من أن هذا لا يعني بالضرورة نقصًا في التماسك الزمني، إلا أنه يوضح قصور CNNs في الاستفادة من المعلومات الزمنية بالشكل الأمثل، وهو ما قد يؤثر سلبًا على جودة الفيديو النهائي.

2.4.1.2– الطرق المعتمدة على الشبكات التكرارية (RNNs)

تعد الشبكات التكرارية (RNNs)، وخاصة شبكات الذاكرة طويلة الأجل (LSTM)، مناسبة للغاية لمعالجة التسلسلات الزمنية مثل الفيديوهات. تتميز RNNs بقدرتها على الاحتفاظ بالمعلومات الزمنية واستيعاب التبعيات الطويلة المدى بين الإطارات.

- FRVSR (Frame-Recurrent Video Super-Resolution : تعتمد FRVSR على بنية تكرارية تستخدم الإطار المحسّن في الخطوة السابقة كمدخل لتحسين الإطار الحالي. يعني ذلك أن الخوارزمية تستفيد من الإطارات السابقة المُعالجة لتوفير معلومات إضافية عند تحسين الإطارات الجديدة، مما يساعد في الحفاظ على التماسك الزمني بين الإطارات المتتابعة. هذا النهج يقلل من التشوهات الزمنية، مثل التموجات أو الاضطرابات التي قد تظهر عند معالجة الإطارات بشكل مستقل. يُظهر نموذج FRVSR فعالية كبيرة في الحفاظ على الحركة الطبيعية والسلسة في الفيديو، مما يجعله مثاليًا للتطبيقات التي تحتاج إلى دقة زمنية عالية، مثل البث المباشر وتحليل الفيديو في الأنظمة الأمنية، حيث يعتبر استقرار الحركة والاتساق الزمني أمراً بالغ الأهمية [11].

- RBPN (Recurrent Back-Projection Network): تجمع هذه الخوارزمية بين الإسقاط الخلفي والشبكات التكرارية لتحسين التفاصيل المكانية والزمنية. على الرغم من أن RBPN تحسن من جودة الفيديوهات المعززة، إلا أن نجاحها يعتمد بشكل كبير على دقة تقدير الحركة، مما قد يؤدي إلى مشاكل إذا كانت الحركة في الفيديو معقدة أو غير متوقعة [12].

سلبيات الشبكات التكرارية (RNNs):

1. التعقيد الحسابي والبطء: الشبكات التكرارية، وخاصة LSTM، تعاني من تعقيد حسابي عالٍ مقارنةً بـ CNNs وعدم القدرة على التنفيذ باستخدام البرمجة التفرعية (التنفيذ على وحدات المعالجة الرسومية GPUs)، مما يؤدي إلى بطء في التدريب والتنفيذ، خاصة عند التعامل مع تسلسلات فيديو طويلة[11].
2. صعوبة التعامل مع التسلسلات الطويلة: تواجه RNNs مشاكل مثل تلاشي التدرجات عندما يتم تدريبها على تسلسلات زمنية طويلة، مما قد يؤدي إلى فقدان المعلومات المهمة المتعلقة بالتبعية الزمنية البعيدة[12].
3. الحساسية لتقدير الحركة: تعتمد جودة الإطارات الناتجة عن RNNs بشكل كبير على دقة تقدير الحركة بين الإطارات، مما يعني أن الأخطاء في تقدير الحركة يمكن أن تؤدي إلى تشوهات ملحوظة[11].

3.4.1.2– الطرق المعتمدة على شبكات الخصوم التوليدية (GANs)

تُعد شبكات الخصوم التوليدية (GANs) من أحدث الأساليب وأكثرها تأثيرًا في تحسين الدقة للفيديو، حيث تتيح توليد تفاصيل دقيقة وواقعية، مما يجعلها ملائمة بشكل كبير لهذا الغرض. كان نموذج **SRGAN**، (Super-Resolution Generative Adversarial Network) الذي طوره Ledig وآخرون [9]، خطوة رائدة في تطبيق GANs لتحسين دقة الصور، ولاحقًا تم توسيع هذا المفهوم ليشمل الفيديو في نماذج مثل نموذج **TecoGAN** (Temporally Coherent Generative Adversarial Network) التي قدمها Chu وآخرون [13].

- **TecoGAN** (**Temporally Coherent Generative Adversarial Network**): تقوم TecoGAN بتوسيع عمل SRGAN من خلال دمج مفهوم التماسك الزمني. يساعد هذا النهج على تقليل التشوهات الزمنية والحفاظ على اتساق التفاصيل عبر الإطارات، مما يؤدي إلى إنتاج فيديوهات أكثر سلاسة واتساقًا بمرور الزمن. يحقق TecoGAN ذلك من خلال استخدام شبكة تلافيفية مسؤولة عن توليد الإطارات ومميز (Discriminator) يتم تدريبه للتحقق من واقعية التفاصيل في كل من الأبعاد المكانية والزمنية.

- **MSRN** (**الشبكة المتبقية متعددة المقاييس**) التي طُورت بواسطة Li وآخرين [24] تجمع بين شبكات الخصومة التوليدية (GANs) وتقنية التعلم المتبقي متعدد المقاييس بهدف تحسين التفاصيل المكانية (مثل الحواف) والزمنية (مثل سلاسة الحركة) في الفيديو، مما يسهم في تحسين أداء تكبير الفيديو بشكل كبير. وعلى الرغم من فعالية هذا النهج في تحسين جودة الفيديو، إلا أن نجاحه يعتمد على إيجاد توازن دقيق أثناء عملية التدريب بين ثلاثة أنواع من توابع الخسارة:

1. تابع الخسارة العدائية (Adversarial Loss): التي تهدف إلى جعل الفيديو الناتج يبدو واقعيًا من خلال استخدام شبكة مميزة (Discriminator) تميز بين الفيديو المعزز والفيديو الحقيقي.
2. تابع الخسارة الإدراكية (Perceptual Loss): التي تُركّز على تحسين التفاصيل الدقيقة بناءً على ما يعتبره البشر مهمًا في جودة الصورة، مثل الحواف والقوام (texture) .
3. تابع خسارة المحتوى (Content Loss) : التي تُستخدم لضمان أن الفيديو الناتج يحافظ على المحتوى الأصلي من حيث الشكل والخصائص.

هذا التوازن ضروري لأنه إذا تم التركيز كثيرًا على دالة الخسارة العدائية، قد يصبح الفيديو غير طبيعي، وإذا تم التركيز فقط على دالة الخسارة الإدراكية، قد يُحسن التفاصيل لكنه يفتقر للواقعية. لذلك، يجب ضبط هذه الخسائر بعناية لضمان إنتاج فيديو بجودة عالية مع تفاصيل دقيقة وواقعية.

سلبيات الشبكات الخصومة التوليدية (GANs) :

1. تحديات التوازن في التدريب :يعد تدريب GANs أمرًا صعبًا بسبب الحاجة إلى تحقيق توازن دقيق بين المولد (Generator) و المميز (Discriminator) إذا لم يتم تحقيق هذا التوازن، فقد يؤدي ذلك إلى مشاكل مثل عدم الاستقرار أثناء التدريب أو إنتاج صور ذات جودة منخفضة [23] .
2. التعقيد الحسابي :تعد GANs من النماذج المعقدة حسابيًا، حيث تتطلب موارد حاسوبية كبيرة للتدريب، مما يجعلها غير ملائمة للاستخدام في البيئات ذات الموارد المحدودة أو في التطبيقات الفورية[13].
3. الحساسية للأنماط الزمنية :على الرغم من أن TecoGAN يعالج التماسك الزمني، إلا أن الحفاظ على الاتساق في المشاهد ذات الحركة السريعة أو المعقدة لا يزال يشكل تحديًا[13].

4.4.1.2– الطرق المعتمدة على نماذج قائمة على المحولات (Transformers)

تم تطوير نماذج المحولات (Transformers) في الأصل لمعالجة اللغات الطبيعية، لكنها أثبتت مؤخرًا فعاليتها في تحسين دقة الفيديو، محققة نتائج واعدة. من بين هذه التطبيقات، تعتمد شبكة المحولات الزمنية المكانية (Spatio-Temporal Transformer Network - STTN)التي اقترحها Zeng وآخرون [18] على آليات الانتباه الذاتي لالتقاط التبعية طويلة المدى عبر كل من الأبعاد المكانية والزمنية، مما يساعد في معالجة التحديات المتعلقة بالحركة المعقدة والتغطيات في

تسلسلات الفيديو. بالإضافة إلى ذلك، ظهرت خوارزمية Swin Transformer كأحد التطورات الحديثة في هذا المجال، حيث تقدم بنية متعددة المستويات تستفيد من النوافذ المنزلقة (Sliding Windows) لتحقيق تكامل أفضل بين المعلومات المكانية والزمنية، مما يعزز دقة استعادة التفاصيل في الفيديوهات المعقدة.[19]
على الرغم من أن المحولات مثل STTN وSwin Transformer تقدم تحسينات ملحوظة في معالجة الحركات المعقدة والتغطيات، إلا أنها قد تواجه صعوبة في التعامل مع الحركات السريعة جدًا أو الحركات التي تتطلب تعويضًا دقيقًا للموضع بين الإطارات المتتالية. هذا يمكن أن يؤدي إلى ظهور تشوهات أو فقدان للاتساق الزمني في الفيديو المعزز [18, 19].

على الرغم من التقدمات الكبيرة في تحسين دقة الفيديو، إلا أن معظم الخوارزميات التي تمت مناقشتها تعتمد بشكل أساسي على الشبكات العصبونية ذات القيم الحقيقية. هذه الشبكات تركز غالبًا على معلومات السعة (Magnitude) أثناء عملية إعادة بناء الفيديو، مما يعني أنها تعطي الأولوية لشدة أو قوة الإشارة في كل بيكسل من الصورة أو الفيديو. في المقابل، تُهمل هذه الشبكات عادةً معلومات الطور(Phase)، التي تحمل تفاصيل مهمة حول البنية الهندسية والترتيب المكاني لعناصر الصورة.

السعة تعكس مدى إضاءة أو لون البيكسل، بينما الطور يُمثل الزاوية النسبية للإشارة، مما يؤثر بشكل كبير على كيفية توزيع التفاصيل الدقيقة، مثل الحواف والأنسجة، عبر الصورة أو الفيديو. بحسب ما وضحه Oppenheim و Lim [20]، فإن الطور يلعب دورًا حاسمًا في الحفاظ على الشكل والترتيب الهندسي للمعلومات البصرية. لذلك، إهمال الطور يمكن أن يؤدي إلى فقدان الكثير من التفاصيل الدقيقة والهيكلية، حتى لو تم إعادة بناء السعة بشكل مثالي.
نتيجة لذلك، اتجهت الأبحاث الحديثة نحو استخدام الشبكات العصبونية ذات القيم العُقدية. هذه الشبكات لديها القدرة على معالجة كل من السعة والطور بشكل متكامل، مما يسمح بتحسين دقة الفيديو بشكل أكثر فعالية من الشبكات التقليدية ذات القيم الحقيقية[15,16].

2.2– الشبكات العصبونية ذات القيم العُقدية (CVNNs)

بينما تركز معظم الخوارزميات الحالية لتحسين الدقة للفيديو على معالجة الإشارات ذات القيم الحقيقية، إلا أن هذا النهج قد يغفل عن بعض المعلومات الهامة المرتبطة بالطور (Phase) والسعة (Magnitude) للإشارة. تمثل الشبكات العصبونية ذات القيم العُقدية (Complex-Valued Neural Networks, CVNNs) تطورًا حديثًا في هذا السياق، حيث تتيح معالجة كل من الطور والسعة معًا، مما يمكن أن يعزز من قدرة الخوارزميات على استعادة التفاصيل الدقيقة وتحقيق التماسك الزمني [15,16].

1.2.2– البنية الرياضية للشبكات العصبونية ذات القيم العُقدية

تعتمد الشبكات العصبونية ذات القيم العُقدية على استخدام أعداد عُقدية في تمثيل المدخلات، الأوزان، والمخرجات.
العدد المركب يتمثل بالشكل التالي:

$$z = x + iy \qquad (1)$$

حيث (x) هو الجزء الحقيقي و(y) هو الجزء التخيلي، و(i) هو الوحدة التخيلية $i^2 = -1$.

2.2.2– العمليات الرياضية الأساسية في CVNNs

– **الضرب المصفوفي المركب**: عند التعامل مع أوزان عُقدية ومدخلات عُقدية، يتم إجراء العمليات الرياضية بطريقة عُقدية. إذا كانت مصفوفة الأوزان العُقدية (W = A + iB) ودخل الشبكة المركب هو (X = C + iD)، فإن ناتج الضرب المصفوفي يُحسب كما يلي:

$$WX = (AC - BD) + i(AD + BC) \qquad (2)$$

هذا النوع من العمليات يسمح للشبكة بالاحتفاظ بمعلومات الطور و السعة معًا خلال عملية التدريب [15].

– **تابع التفعيل المركب**: في CVNNs، يتم تطبيق تابع التفعيل على كل من الجزء الحقيقي والجزء التخيلي بشكل مستقل. على سبيل المثال، إذا كان تابع التفعيل هو ReLU، يتم تطبيقه على الجزئين كلاً على حدة:

$$f(z) = ReLU(x) + iReLU(y) \qquad (3)$$

هذا يضمن معالجة غير خطية للإشارات العُقدية، مما يعزز من قدرة الشبكة على التكيف مع الأنماط المعقدة [16].

3.2.2– أهمية CVNNs في تحسين دقة الفيديو

تتيح CVNNs معالجة شاملة لكل من الطور والسعة في إشارة الفيديو، مما يجعلها مفيدة بشكل خاص في التطبيقات التي تتطلب استعادة دقيقة للتفاصيل الهيكلية والحفاظ على التماسك الزمني. من خلال الاستفادة من القدرات الفريدة لـ CVNNs، يمكن تحسين جودة الفيديوهات الناتجة بشكل ملحوظ مقارنة بالشبكات العصبونية التقليدية ذات القيم الحقيقية، خاصة في سيناريوهات الحركة المعقدة أو التفاصيل الدقيقة [15,16].

4.2.2– أهمية الطور في تحسين دقة الفيديو:

يلعب الطور دورًا حيويًا في تحسين دقة الفيديو، لأنه يحدد كيفية ترتيب الترددات العالية، مثل الحواف والتفاصيل الدقيقة، داخل الإطار. بينما تعبر الترددات العالية عن تفاصيل الصورة الدقيقة، يُحدد الطور كيف يتم توزيع هذه التفاصيل في المكان. إهمال الطور، كما يحدث في الشبكات ذات القيم الحقيقية، قد يؤدي إلى تشوهات في ترتيب هذه التفاصيل، مما يؤثر سلبًا على جودة الفيديو. باستخدام الشبكات العصبونية ذات القيم العُقدية، التي تعالج الطور والسعة معًا، يمكن الحفاظ على كل من شدة وتوزيع التفاصيل الدقيقة والحواف، مما يؤدي إلى تحسين شامل في جودة الفيديو النهائي[17] .

3.2– مقاييس التقييم ومجموعات البيانات

1.3.2– مقاييس التقييم

يتم تقييم أداء نماذج تحسين الدقة الفائقة للفيديو (video super-resolution) باستخدام عدة مقاييس كمية ونوعية، كل منها يقدم نظرة مختلفة على جودة الفيديو الناتج مقارنة بالفيديو الأصلي عالي الدقة. تشمل المقاييس الرئيسية:

1.1.3.2– نسبة الإشارة إلى الضوضاء (**PSNR – Peak Signal-to-Noise Ratio**) :

يعد PSNR من أكثر المقاييس الكمية شيوعًا لتقييم جودة الصور والفيديو[7]. يقيس PSNR مدى دقة الفيديو المعاد بناؤه مقارنةً بالفيديو الأصلي عالي الدقة (Ground Truth) عن طريق حساب نسبة الإشارة (المعلومات المفيدة) إلى الضوضاء (التشوهات) في الإشارة المعاد بناؤها.

-حساب PSNR: يعتمد على حساب الفرق بين قيم البيكسل في الفيديو المعاد بناؤه والفيديو الأصلي باستخدام مقياس الخطأ المتوسط التربيعي (MSE). يُحسب PSNR من خلال المعادلة:

$$PSNR = 10\log_{10}\left(\frac{L^2}{MSE}\right) \qquad \textbf{(4)}$$

حيث L هو أقصى قيمة ممكنة للبيكسل (عادةً 255)، و MSE هو متوسط مربع الخطأ بين الفيديو المعاد بناؤه والفيديو الأصلي و يحسب كالتالي :

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{I}_i - \tilde{I}_i\right)^2 \qquad \textbf{(5)}$$

في هذه المعادلة:

- **N** هو العدد الكلي للبيكسلات في الصورة أو الإطار.
- $\hat{I}$ هو إطار الفيديو الأصلي عالي الدقة.(HR)
- $\tilde{I}$ هو الإطار المعاد بناؤه بالدقة الفائقة.(SR)
-

يشير PSNR الأعلى إلى أن الفيديو المعاد بناؤه أقرب إلى الفيديو الأصلي من حيث دقة البيكسلات، وبالتالي إلى جودة أفضل.

2.1.3.2– مؤشر التشابه الهيكلي (**SSIM – Structural Similarity Index Measure**)

SSIM هو مقياس يعتمد على تقييم التشابه الهيكلي بين الفيديو المعاد بناؤه والفيديو الأصلي، ويعتبر أكثر توافقًا مع الإدراك البشري مقارنةً بـ PSNR.

يأخذ SSIM بعين الاعتبار ثلاثة عوامل رئيسية – السطوع، التباين، والبنية. يتم حساب التشابه بين الفيديوهات الأصلية والمعاد بناؤها عبر تحليل كل جزء من الصورة بدلاً من تقييم الصورة ككل. يعتمد على قياس التوافق الهيكلي، حيث يتم تحليل الصور اعتمادًا على تشابه الأنماط الهيكلية بين الفيديو المعاد بناؤه والفيديو الأصلي.[10]

يُحسب SSIM من خلال المعادلة:

$$SSIM\ (I^{\wedge}, I^{\sim}) = \frac{2u_{\hat{I}}u_{\tilde{I}} + k_1}{u_{\hat{I}}^2 + u_{\tilde{I}}^2 + k_1} \cdot \frac{2\sigma_{\hat{I}\tilde{I}} + k_2}{\sigma_{\hat{I}}^2 + \sigma_{\tilde{I}}^2 + k_2} \qquad (6)$$

حيث:

- $u_{\tilde{I}}$, $u_{\hat{I}}$: هما متوسطا قيم الصور $I^{\wedge}, I^{\sim}$على التوالي.
- $\sigma_{\tilde{I}}$, $\sigma_{\hat{I}}$: هما الانحرافان المعياريان للصورتين.
- $\sigma_{\hat{I}\tilde{I}}$: هو التباين المشترك بين الصورتين.
- k_1, k_2 : هما ثوابت تُستخدم لتثبيت الحساب، وعادةً ما تُضبط على القيم 0.01 و 0.03 على التوالي.

يشير SSIM الأعلى إلى جودة أفضل وإدراك مرئي أفضل، حيث يُظهر مدى التشابه بين البنية العامة للفيديو المعاد بناؤه والفيديو الأصلي.

3.1.3.2– تشابه الرقعة الصورية الإدراكية المتعلمة (LPIPS – Learned Perceptual Image Patch Similarity)

LPIPS هو مقياس إدراكي حديث يقيم التشابه المرئي بين الفيديوهات باستخدام الميزات العميقة من الشبكات العصبونية المدربة مسبقًا. تم تقديم هذا المقياس بواسطة Zhang وآخرين [11]، وهو يعتمد على التمثيلات المتعلمة للصور من خلال شبكات عصبونية تلافيفية عميقة (CNNs) مدربة على التعرف على الصور.

يتم استخراج الميزات العميقة من شبكات CNN مثل VGG أو AlexNet، ثم يُحسب التشابه بين هذه الميزات في الفيديو المعاد بناؤه والفيديو الأصلي. يُستخدم هذا المقياس لتحليل تفاصيل دقيقة يصعب إدراكها من خلال المقاييس التقليدية مثل PSNR وSSIM، حيث يعتمد على كيفية إدراك البشر للتفاصيل والتشابهات.

يشير LPIPS المنخفض إلى تشابه بصري أكبر بين الفيديو المعاد بناؤه والفيديو الأصلي، مما يعني أن الفيديو الناتج ذو جودة إدراكية عالية وقريبة من الجودة الأصلية. يُستخدم هذا المقياس بشكل متزايد لكونه يتوافق مع الإدراك البشري بشكل أفضل من PSNR وSSIM، خاصةً في المهام التي تتطلب تفاصيل دقيقة وجودة عالية.

2.3.2– مجموعات البيانات المرجعية

تعتمد أبحاث تحسين الدقة الفائقة للفيديو على عدة مجموعات بيانات مرجعية متنوعة، تستخدم لتدريب وتقييم أداء النماذج. توفر هذه المجموعات مشاهد واقعية تتضمن تحديات حقيقية، مما يسمح للنماذج باختبار قدرتها على استعادة التفاصيل الدقيقة وتحقيق جودة بصرية عالية. فيما يلي أبرز مجموعات البيانات المستخدمة:

- **Vid4**: تُعتبر Vid4واحدة من أقدم مجموعات البيانات المستخدمة في أبحاث تحسين دقة الفيديو. تم إنشاؤها واستخدامها لأول مرة في أبحاث[7] . و تحتوي هذه المجموعة على أربعة تسلسلات فيديو ، حيث توفر كل سلسلة أنماطًا حركية مختلفة، مثل الحركات السلسة والبطيئة والحركات السريعة والمفاجئة. يصل عدد الإطارات الإجمالي في المجموعة إلى 171 إطارًا فقط، مما يجعلها صغيرة نسبيًا.

  على الرغم من صغر حجمها، تظل Vid4 معيارًا مرجعيًا شائعًا للاختبارات الأولية، حيث كانت تستخدم على نطاق واسع في الأبحاث التقليدية التي تهدف إلى تقييم النماذج في بيئات محدودة الحجم قبل الانتقال إلى المجموعات الأكبر.

- **REDS** (**مشاهد واقعية ومتنوعة**): تم تطوير مجموعة بيانات REDSكجزء من تحدي NTIRE 2019 لتحسين دقة الفيديو وإزالة الضبابية، كما ورد في بحث[8] . تحتوي REDS على 239تسلسل فيديو، يتراوح طول كل منها بين 100 و120 إطارًا، مما يجعلها مجموعة بيانات كبيرة الحجم وتتميز بواقعية عالية.

  تقدم REDS مجموعة متنوعة من المشاهد اليومية، تشمل مشاهد داخلية، خارجية، ومشاهد في الشوارع، حيث توفر ظروفًا متباينة مثل الإضاءة المنخفضة، وضبابية الحركة الواقعية، وتباين الإضاءة، ما يتيح للنماذج اختبار قدرتها على استعادة التفاصيل الدقيقة في ظروف متنوعة وواقعية. ونظرًا للتنوع الكبير في هذه المجموعة، تعد REDS مناسبة بشكل خاص لتقييم أداء النماذج المتقدمة التي تهدف إلى تحقيق تحسينات كبيرة في جودة الفيديو ضمن سياقات حركية معقدة.

- **Vimeo-90K**: تعد Vimeo-90Kمن أكبر مجموعات البيانات المستخدمة في أبحاث تحسين دقة الفيديو وأكثرها شمولاً، وقد تم تقديمها من قبل[9] . تتكون المجموعة من حوالي 90,000مقطع فيديو قصير تم استخراجها من منصة Vimeo، مما يوفر مجموعة واسعة من المشاهد الحركية والإضاءة المتنوعة.

  تحتوي Vimeo-90K على تسلسلات فيديو مقسمة إلى تسع فئات مختلفة من الحركة والإضاءة، مما يسهم في تقييم قدرة النماذج على التكيف مع مجموعة واسعة من الظروف. تتميز مقاطع الفيديو بتسلسل من 7 إطارات عالية الدقة، مما يوفر للنماذج سياقًا زمنيًا متعدد الإطارات يساعدها على تعلم الميزات الزمنية والمكانية بشكل فعال. تعتبر هذه المجموعة معيارًا مرجعيًا هامًا في الأبحاث الحديثة، حيث تُستخدم على نطاق واسع لتدريب النماذج وتقييمها في مهام تتطلب دقة إدراكية عالية.

الفصل الثالث

النموذج المقترح

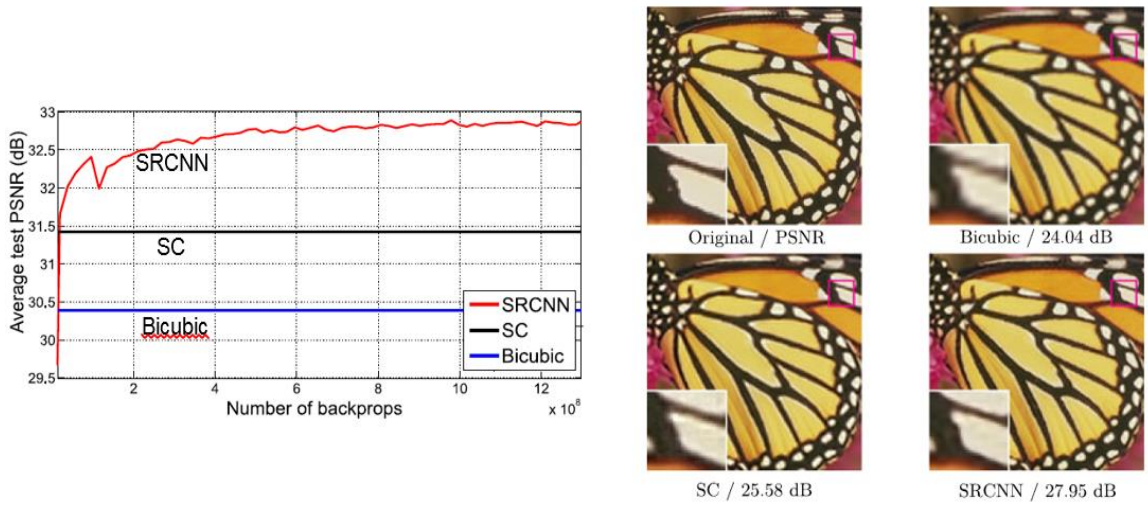
في هذا الفصل، سيتم استعراض الخوارزميات الأساسية المستخدمة في البحث وتوضيح الخطوات المنهجية المتبعة لتحسين دقة الفيديو باستخدام شبكات عصبية ذات قيم عُقدية كحل مقترح لمسألة البحث.

يهدف هذا الفصل إلى تحليل ودراسة ثلاث خوارزميات رائدة في تحسين دقة الفيديو، وهي SRCNN و FRVSR وTecoGAN، حيث سيتم استعراض بنية كل خوارزمية وآلية عملها، مع التركيز على كيفية تحويل هذه الخوارزميات إلى شبكات عصبية ذات قيم عُقدية. يسعى هذا التحويل إلى تعزيز قدرة النماذج على استعادة التفاصيل بشكل أكثر كفاءة ودقة. سنبدأ بشرح خوارزمية SRCNN، التي تُعد من أوائل الخوارزميات التي اعتمدت على الشبكات العصبونية التلافيفية (CNN) لتحسين دقة الصور، مما وضع الأساس للتطورات اللاحقة في تحسين الدقة للفيديو.

1.3– الخوارزمية SRCNN (Super Resolution Convolutional Neural Network)

كانت خوارزمية SRCNN (Super Resolution Convolutional Neural Network) التي طُرحت في البحث [21]من أوائل الأعمال الرائدة المعتمدة على التعلم العميق في مجال الدقة الفائقة، وتشكل الأساس الذي بُنيت عليه معظم الخوارزميات اللاحقة المعتمدة على التعلم العميق في هذا المجال. وعلى الرغم من بساطتها، فقد أظهرت SRCNN تفوقًا ملحوظًا من حيث جودة الصور عالية الدقة مقارنة بأفضل الطرق المعتمدة على الأمثلة مثل طريقة الترميز المتناثر SC [5] (كما هو موضح في الشكل 6)، بالإضافة إلى تفوقها في سرعة التنفيذ على وحدة المعالجة المركزية(CPU) .

تجدر الإشارة إلى أن أداء الخوارزميات المعتمدة على التعلم العميق يتحسن بشكل كبير عند استخدامها لمجموعة تدريب كبيرة ومتنوعة من الصور، إذ تتيح هذه المجموعات الغنية تحسينات في جودة الصور المعاد بناؤها بدقة عالية. غير أن مثل هذه المجموعات الكبيرة تشكل تحديًا بالنسبة للأساليب المعتمدة على الأمثلة، نظرًا لصعوبة تحقيق عمليات الربط الصريح (explicit) والمثلى (optimized)بين الصور[6]. تعتمد خوارزمية SRCNN على شبكة عصبية تلافيفية (CNN) لبناء الصور عالية الدقة بشكل مباشر وشامل(end-to-end learning) ، حيث تقوم هذه الشبكة بتعلم العلاقات المثلى بين الصور منخفضة وعالية الدقة مع الحاجة إلى حد أدنى من المعالجة الأولية واللاحقة (Pre/post-processing) للصور[21].



الشكل 1: مقارنة الخوارزمية SRCNN مع طريقة الاستيفاء التكعيبي وطريقة الترميز المتناثر SC

1.1.3– بنية خوارزمية SRCNN

تنطلق خوارزمية SRCNN من صورة منخفضة الدقة Y ، حيث تقوم أولاً بزيادة حجم هذه الصورة (Upscaling) باستخدام طريقة الاستيفاء التكعيبي، وهي المعالجة الأولية الوحيدة المطلوبة في الخوارزمية. بعد عملية الاستيفاء، تأخذ الصورة Y أبعاد الصورة عالية الدقة المطلوبة، بحيث تسعى الخوارزمية أثناء عملية التعلم إلى إنتاج صورة عالية الدقة انطلاقًا من الصورة Y ، بحيث تكون قريبة قدر الإمكان من الصورة الأصلية عالية الدقة X (Ground truth) الموجودة ضمن مجموعة التدريب، مما يؤدي إلى بناء تابع تقابل يُعبر عنه بالعلاقة $X = F(Y)$. يمكن تقسيم هذه العملية إلى ثلاث مراحل أساسية:

1. **استخراج البقع وتمثيلها (Patch Extraction and Representation)**:

تتضمن المرحلة الأولى استخراج بقع صغيرة من الصورة وتمثيلها باستخدام قواعد فضاءات التمثيل المحددة مسبقًا، (مثل PCA أو DCT أو Haar), والتي تعتبر استراتيجية شائعة في تقنيات استعادة الصور (Image Restoration). تعادل هذه العملية ضرب الصورة بمجموعة من المرشحات، حيث يعمل كل مرشح كقاعدة تمثيلية. تدمج خوارزمية SRCNN عملية تحسين هذه القواعد ضمن تحسين الشبكة نفسها، وتُعبّر الطبقة الأولى من الشبكة عن هذه العملية بواسطة العلاقة التالية:

$$F_1(Y) = max(0, W_1 * Y + B_1) \qquad (7)$$

حيث تمثل W_1 المرشحات (Filters)، و B_1 تمثل الإزاحات (Biases)، بينما تشير العملية * إلى جداء التلاف (Convolution). يتكون W_1 من n_1 مرشحًا بأبعاد $c \times f_1 \times f_1$، حيث c هو عدد القنوات اللونية في الصورة (مثلًا

c = 3 للصور الملونة) و f_1 هي الأبعاد المكانية للمرشح (مثلاً 5×5 بيكسل). بناءً على ذلك، يتم تطبيق n_1 عملية تلاف بأبعاد $c \times f_1 \times f_1$ على صورة الدخل منخفضة الدقة.

يتمثل ناتج هذه المرحلة في n_1 خريطة من السمات (Feature Maps)، حيث يتكون B_1 من شعاع طوله n_1 بحيث تتطابق كل قيمة مع أحد المرشحات المستخدمة. في هذه المرحلة، يتم تطبيق تابع التفعيل ReLU (Rectified Linear Unit) والتي تعرف بالعلاقة ReLU =max$(0, x)$ على استجابة كل مرشح من المرشحات المستخدمة في W_1.

2. عملية التقابل اللاخطية (Non-linear mapping):

في الطبقة الأولى من الشبكة، يتم تمثيل كل بقعة مستخرجة من الصورة بشعاع سمات طوله n_1 وفي المرحلة الثانية، يتم تحويل كل شعاع سمات من الطول n_1 إلى شعاع جديد بطول n_2 تُعادل هذه العملية تطبيق n_2 مرشحًا بأبعاد 1×1، والتي تمثل الطبقة الثانية من الشبكة وتُعبر عنها بالعلاقة التالية:

$$F_2(Y) = max(0, W_2 * F_1(Y) + B_2) \qquad \textbf{(8)}$$

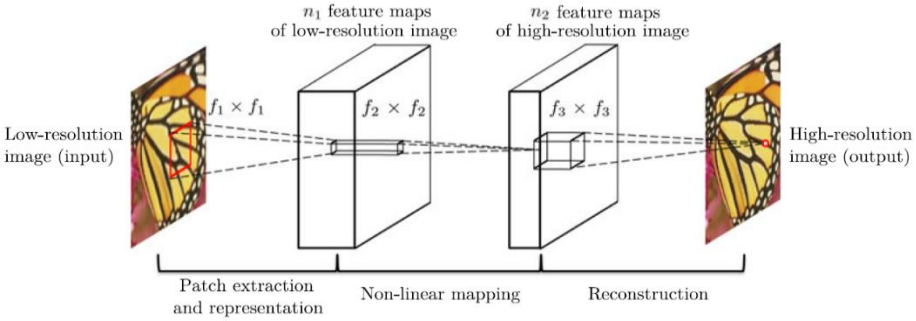
حيث يحتوي W_2 على n_2 مرشحًا، كل منها بأبعاد $n_1 \times f_2 \times f_2$، بينما B_2 هو شعاع بطول n_2.يتمثل ناتج هذه الطبقة في أشعة سمات جديدة بطول n_2، يمثل كل منها بقعة ذات دقة عالية. سيتم استخدام هذه البقع في الخطوات التالية لبناء الصورة عالية الدقة.

3. إعادة البناء (Reconstruction):

في الأساليب التقليدية، يتم غالبًا حساب متوسط (Averaging) البقع عالية الدقة المتداخلة الناتجة من أجل إعادة بناء الصورة النهائية، وهذه العملية تعادل تطبيق مرشح توسيط على مجموعة خرائط السمات. بناءً على هذا المبدأ، يمكن التعبير عن الطبقة الثالثة من الشبكة SRCNN بالمعادلة التالية:

$$F(Y) = W_3 * F_2(Y) + B_3 \qquad \textbf{(9)}$$

حيث W_3 هو مجموعة من المرشحات عددها c(عدد أقنية الصورة) وكل منها بأبعاد $n_2 \times f_3 \times f_3$ بينما B_3 هو شعاع بطول c. يوضح الشكل 2 المراحل الثلاث للخوارزمية SRCNN.



الشكل 2: المراحل الثلاث للخوارزمية SRCNN

2.1.3– تدريب الشبكة

تعتمد عملية التعلم الشامل (End-to-End Learning) لتابع التقابل F (أو تدريب الشبكة) في خوارزمية SRCNN على تقدير مجموعة المعاملات $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$, ويتم ذلك بتصغير تابع الكلفة (Loss Function) بين الصورة عالية الدقة المتوقعة $F(Y, \Theta)$والصورة الأصلية X المقابلة (Ground Truth). باستخدام مجموعة صور منخفضة الدقة $\{Y_i\}$وصورها عالية الدقة $\{X_i\}$، يُستخدم متوسط الخطأ التربيعي MSE (Mean squared error) كتابع كلفة:

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\left\|F(Y_i, \Theta) - X_i\right\|^2 \qquad \textbf{(10)}$$

حيث n هو عدد الصور في التدريب. يتم تصغير MSE باستخدام الانحدار التدريجي العشوائي (Stochastic gradient descent) مع الانتشار الخلفي المعياري (Standard backpropagation):

$$\Delta_{i+1} = 0.9\Delta_i + \eta\frac{\partial L}{\partial W_i^l} \;\; ; \;\; W_{i+1}^l = W_i^l + \Delta_{i+1} \qquad \textbf{(11)}$$

حيث $l \in \{1, 2, 3\}$ هو رقم الطبقة ضمن الشبكة العصبونية (الشبكة المقترحة مكونة من ثلاث طبقات مخفية فقط)، η هو معدل التعلم (Learning rate). يتم تهيئة الأوزان $\{W_0^l\}$ لكل مرشح في كل طبقة باستخدام توزيع غوصي (Gaussian distribution) بوسطي معدوم (Zero mean) وانحراف معياري 0.001 أما بالنسبة للإنزياحات $\{B_0^l\}$ فتعطى قيم صفرية. بالنسبة لمعدل التعلم فيعطى قيمة 10^{-4} بالنسبة للطبقتين الأولى والثانية وقيمة 10^{-5} بالنسبة للطبقة الثالثة حيث تبين تجريبياً أن تصغير معدل التعلم في الطبقة الثالثة ضروري لتقارب الشبكة العصبونية. لابد من الإشارة هنا إلى أن مجموعة التدريب غالباً ما تكون مكونةً فقط من مجموعة من الصور عالية الدقة ولذلك أثناء مرحلة التعلم يتم الحصول على الصور

منخفضة الدقة انطلاقاً من الصور عالية الدقة. تمت هذه العملية ضمن الخوارزمية المقترحة من خلال إضافة عدم وضوح بصري (Blur) للصورة عالية الدقة باستخدام نواة غوصية (Gaussian kernel) ومن ثم تصغير هذه الصورة (Downscaling) بنفس معامل التكبير (Upscaling factor) المطلوب، ومن ثم إعادة تكبير هذه الصورة (Upscaling) لنفس الحجم (نفس معامل التكبير) باستخدام طريقة الاستيفاء التكعيبي. لقد تم التخلص من أثر الحواف أثناء عملية التدريب من خلال عدم إضافة أصفار (Padding) لحواف الصورة، وبالتالي الحصول في الخرج على صورة أصغر بأبعاد $((f_{in} - f_1 - f_2 - f_3 + 3)^2 \times c)$ حيث f_{in} هو عرض (Width) أو ارتفاع (Height) صورة الدخل والذين يعتبران متساويين في الخوارزمية المقترحة (أي صورة الدخل مربعة كما هو الحال في معظم التقنيات المعتمدة على الشبكات العصبونية التلفيفية).

بينما تعد خوارزمية SRCNN فعّالة في تحسين دقة الصور الثابتة، فإنها مصممة خصيصًا لهذا النوع من البيانات ولا تأخذ في الاعتبار العلاقات الزمنية بين الإطارات المتتالية في الفيديو. لذلك، عند تطبيقها على الفيديو، قد تفتقر إلى التماسك الزمني المطلوب لضمان استمرارية التفاصيل بين الإطارات، خاصةً في المشاهد التي تتضمن حركة. على عكسSRCNN ، فإن خوارزمية FRVSR (Frame-Recurrent Video Super-Resolution)[11]تم تطويرها خصيصًا لتحسين دقة الفيديو، حيث تعتمد على إطار مرجعي سابق (الإطار عالي الدقة الذي تم توليده في الخطوة السابقة) وتستفيد من التتابع الزمني للإطارات لتحسين الدقة التفاعلية للإطارات اللاحقة. هذا النهج يتيح الحفاظ على التناسق الزمني واستمرارية التفاصيل عبر مقاطع الفيديو بشكل أفضل.

فيما يلي، سيتم شرح خوارزمية FRVSR بالتفصيل، مع التركيز على بنيتها وآلية عملها في تحسين دقة الفيديو.

2.3- FRVSR (Frame-Recurrent Video Super-Resolution)

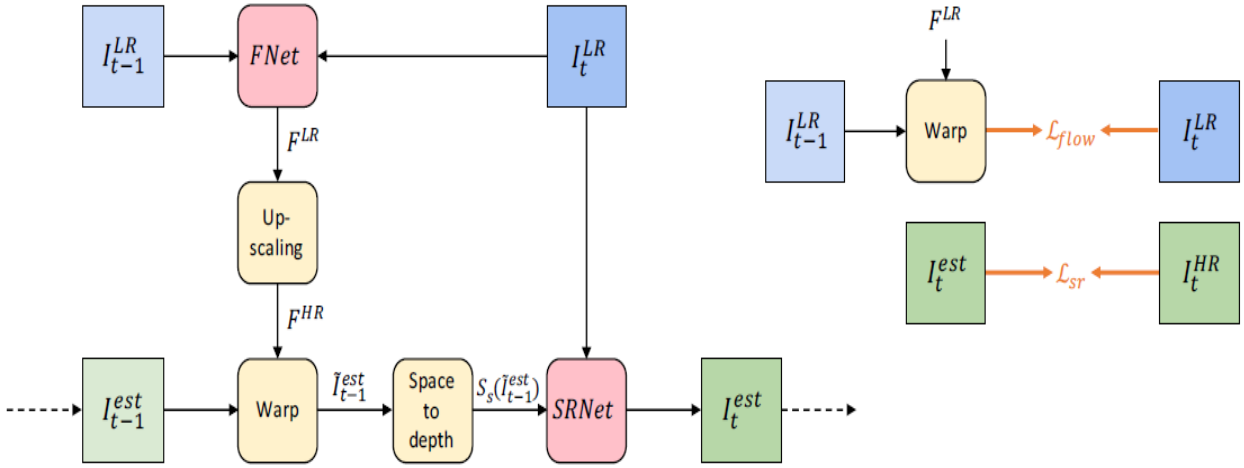
تعتمد خوارزمية **FRVSR** [11]على استخدام الشبكات العصبونية العميقة لتحسين دقة الفيديو، حيث تُستخدم الإطارات عالية الدقة (HR) التي يتم توليدها من إطارات منخفضة الدقة (LR) عبر تقنية **التدفق التكراري** .تقوم الفكرة الأساسية لتقنية التدفق التكراري على استخدام الإطار عالي الدقة الذي تم تحسينه في الخطوة السابقة كمدخل للإطار التالي. يتيح ذلك للخوارزمية استمرارية التفاصيل الزمنية بين الإطارات المتتابعة، ويضمن التناسق الزمني مع تعزيز التفاصيل المكانية. علاوةً على ذلك، يساهم هذا الأسلوب التكراري في تقليل الكلفة الحسابية المطلوبة لمعالجة عدد كبير من الإطارات في الفيديو، مما يسمح بتحسين الجودة الإجمالية دون التأثير على كفاءة الأداء.

2.2.3–البنية العامة لخوارزمية FRVSR:

تتكون خوارزمية FRVSR من مكونين أساسيين (Sajjadi et al., 2018):

1. **شبكة التدفق (FNet)**: مسؤولة عن تقدير الحركة بين الإطارات المتتالية عبر استخراج خريطة حركة بين الإطارات منخفضة الدقة.
2. **شبكة تحسين الدقة (SRNet)** : مسؤولة عن إنتاج الإطار عالي الدقة الجديد باستخدام المعلومات المستخرجة من الإطار الحالي منخفض الدقة والإطار السابق عالي الدقة الذي تم مواءمته عبر خريطة الحركة.

يُظهر الشكل أدناه التدفق الكامل للخوارزمية، بدءًا من تقدير الحركة باستخدام شبكةFNet ، مرورًا بعملية التشويه(Warp) ، ثم تحويل المعلومات لتمكين شبكة SRNet من إنتاج الإطار الجديد عالي الدقة.



الشكل:3 مخطط تدفق خوارزمية FRVSR موضحًا شبكتي FNet وSRNet

3.2.3– الآلية التفصيلية لعمل FRVSR:

1. تقدير الحركة باستخدام شبكة التدفق البصري FNet (Flow Estimation)

في البداية، تقوم الخوارزمية بتقدير التدفق البصري (Optical flow) بين الإطارات المتتابعة منخفضة الدقة I_{t-1}^{LR} و I_t^{LR} يتم ذلك باستخدام شبكة تدفق البصري $FNet$ ، التي تنتج خريطة تدفق بصرية منخفضة الدقة F^{LR}، والتي تعبر عن (الاتجاه و المسافة)كيفية انتقال كل بكسل من الإطار السابق إلى الإطار الحالي.

$$F^{LR} = FNet(I_{t-1}^{LR}, I_t^{LR}) \quad (12)$$

حيث $F^{LR} \in [-1,1]^{\{H \times W \times 2\}}$ تمثل خريطة التدفق البصري بين الإطارات منخفضة الدقة والتي تتكون من قناتين الأولى هي تمثل الإزاحة الأفقية(Horizontal displacement) والثانية هي تمثل الإزاحة الرأسية (Vertical displacement)

2. تكبير خريطة التدفق (Upscaling Flow)

بعد الحصول على خريطة التدفق البصري منخفضة الدقة F^{LR} نقوم بتكبيرها الى الدقة العالية F^{HR} بإستخدام الاستيفاء الثنائي (bilinear interpolation) لملاءمة خريطة التدفق للمعالجة على مستوى الإطار عالي الدقة.

$$F^{HR} = Up(F^{LR}) \quad (13)$$

حيث $F^{HR} \in [-1,1]^{\{H_s \times W_s \times 2\}}$ خريطة التدفق البصري المكبرة إلى الدقة العالية.

3.تراكب الإطار السابق(Warping the Previous Output)

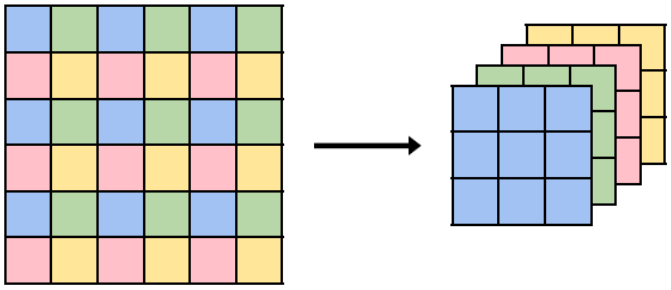
باستخدام خريطة التدفق البصري عالية الدقة F^{HR}، يتم تراكب التقدير السابق للإطار عالي الدقة I_{t-1}^{est} مع الإطار الحالي باستخدام تابع التراكب WP ، ما يسمح بإعادة استخدام معلومات التفاصيل من الإطار السابق.

$$\tilde{I}_{t-1}^{est} = WP(I_{t-1}^{est}, F^{HR}) \quad (14)$$

4.التحويل من الفضاء إلى العمق (Mapping to LR Space via Space-to-Depth Transformation)

الخطوة التالية هي تحويل الصورة التقديرية السابقة $\tilde{I}_{t-1}^{est}$ إلى شبكة منخفضة الدقة باستخدام عملية **التحويل من الفضاء إلى العمق** S_s حيث يتم تقسيم الصورة عالية الدقة إلى شبكة منخفضة الدقة مع قنوات متعددة.

شرح التحويل من الفضاء إلى العمق



الشكل:4 التحويل من الفضاء إلى العمق،

في الشكل المرفق، توضح العملية كيفية أخذ صورة عالية الدقة وتقسيمها إلى شبكات منخفضة الدقة بأبعاد متساوية ولكن ذات إزاحات مختلفة.

يمثل التحويل من الفضاء إلى العمق (**Space–to–Depth**)انتقال البيانات من صورة واحدة عالية الدقة إلى عدة شبكات منخفضة الدقة متراكبة، حيث تنتقل كل منطقة من الصورة الأصلية إلى قناة مختلفة ضمن الشبكة الناتجة. يوضح هذا التحويل رياضيًا على النحو التالي:

$$S_s: [0,1]^{\{sH \times sW \times C\}} \rightarrow [0,1]^{\{H \times W \times s^2 C\}}$$

وهذا التحويل يتم تطبيقه كالآتي:

$$S_s(I)_{\{i,j,k\}} = I_{\left\{si + k \,\%\, s, sj + \left(\frac{k}{s}\right) \%\, s, \frac{k}{s^2}\right\}} \qquad \textbf{(15)}$$

حيث : i,j,k تمثل مؤشرات الشبكة.

$\%$: هو باقي القسمة.

s : هو معامل التحويل الذي يحدد حجم التقسيمات.

يتمثل الغرض الأساسي من هذا التحويل في تقليل كمية البيانات ومعالجة الصورة بكفاءة أكبر دون فقدان الكثير من التفاصيل

5. الإطار النهائي (Super–Resolution)

في الخطوة الأخيرة من خوارزمية **FRVSR**، يتم دمج الإطار منخفض الدقة الحالي I_t^{LR} مع ناتج التحويل من الفضاء إلى العمق للإطار التقديري السابق $S_{s(\tilde{I}_{t-1}^{est})}$ وذلك عبر عملية **دمج قنوي** (concatenation)على بعد القنوات. بعد ذلك، يتم تمرير هذا الناتج إلى **شبكة تحسين الدقة SRNet** ، التي تقوم بمعالجة البيانات وإنتاج الإطار عالي الدقة الجديد I_t^{est}

توضح المعادلة التالية عملية التقدير النهائي للإطار عالي الدقة:

$$I_t^{est} = SRNet(I_t^{LR} \oplus S_{s(\tilde{I}_{t-1}^{est})}) \qquad \textbf{(16)}$$

حيث : $\oplus$ تمثل عملية الدمج القنوي، التي تجمع المعلومات من الإطار الحالي منخفض الدقة والإطار السابق عالي الدقة.

المعادلة النهائية للإطار التقديري فيFRVSR

لإنتاج الإطار عالي الدقة في FRVSR، يتم تجميع جميع العمليات السابقة في صيغة رياضية نهائية، تدمج التقدير الحالي للإطار مع التدفق الحركي بين الإطارات المتتالية. تعمل هذه الصيغة على الجمع بين المعلومات المكانية من الإطار الحالي والمعلومات الزمنية من الإطار السابق وذلك من خلال الدمج مع خريطة التدفق الحركي لتوجيه عملية التقدير.

المعادلة النهائية للتقدير هي:

$$I_t^{est} = \mathrm{SRNet}(\, I_t^{LR} \oplus \mathrm{S_s}(\, \mathrm{WP}(\, I_{t-1}^{est}, \mathrm{UP}(\, \mathrm{FNet}(\, I_{t-1}^{LR}, I_t^{LR}))))) \qquad \textbf{(17)}$$

تساعد هذه العمليات التكرارية المتكاملة في إنتاج إطارات عالية الدقة متسقة مع الإطارات السابقة، مما يضمن تحسين التفاصيل المكانية والزمنية في الفيديو بشكل مستمر مع تقليل الكلفة الحسابية.

4.2.3– تدريب الشبكة و توابع الخسارة

تعتمد عملية تدريب النموذج على استخدام تابعي فقد رئيسيين يهدفان إلى تحسين أداء الشبكة وضبط جودة الإطارات المتوقعة. يتمثل الهدف الأساسي من هذين التابعين في تعزيز دقة الإطارات عالية الدقة وتحسين تقدير التدفق البصري بين الإطارات المتتالية.

- **تابع الخسارة SR (L_{sr})** :

يُستخدم تابع L_{sr} لتقييم جودة الإطار عالي الدقة الناتج من شبكة SRNet، حيث يحسب الفارق التربيعي المتوسط بين الإطار الناتج والإطار المرجعي الحقيقي عالي الدقة.يُعبر عن التابع رياضياً بالصيغة التالية :

$$L_{sr} = || I_t^{est} - I_t^{HR}||_2^2 \quad (18)$$

حيث :

I_t^{est} : يمثل الإطار المقدر عالي الدقة الذي تنتجه الشبكة.

I_t^{HR} : يمثل الإطار المرجعي عالي الدقة.

يتم استخدام L_{sr}لتوجيه عملية التدريب من خلال الانتشار الخلفي (backpropagation) لتحديث أوزان الشبكة. يُطبّق هذا التحديث على كل من شبكة SRNet وشبكة FNet لضمان الحصول على جودة عالية في الإطارات المنتجة.

- **تابع خسارة التدفق (L_{flow})** :

تم تصميم تابع L_{flow} لدعم عملية تقدير التدفق البصري بواسطة شبكة FNet، وذلك نظرًا لعدم توفر بيانات تدفق بصري حقيقية (ground truth) في مجموعة البيانات المستخدمة.

يتم حساب هذا التابع عبر إيجاد متوسط الخطأ التربيعي بين الإطارات منخفضة الدقة المتراكبة عبر خريطة التدفق، مما يسهم في تحسين توافق الشبكة مع الحركة المكانية في الفيديو.تُعبر الصيغة الرياضية لتابع L_{flow} عن هذا الفقد كما يلي:

$$L_{flow} = || WP(I_{t-1}^{LR}, F^{LR}) - I_t^{LR}||_2^2 \quad (19)$$

حيث:

- **WP**: **تشير** إلى عملية التراكب (warping)التي تستخدم خريطة التدفق F^{LR}لتراكب الإطار السابق منخفض الدقة I_{t-1}^{LR} ومحاذاته مع الإطار الحالي.
- I_t^{LR} : يمثل الإطار الحالي منخفض الدقة.

- **تابع الخسارة الكلية (Total Loss)**

يتم الجمع بين تابعي الفقد L_{sr}و L_{flow}للحصول على تابع الفقد الكلي L، والذي يعبر عنه بالمعادلة التالية:

$$L = L_{sr} + L_{flow} \quad (20)$$

يُستخدم تابع الفقد الكلي L خلال عملية التدريب لتحديث أوزان الشبكة، حيث يسهم كل من L_{sr}و L_{flow}في تحسين أداء النموذج بشكل متكامل؛ إذ يركز L_{sr} على تحسين جودة الإطارات المتوقعة، بينما يركز L_{flow} على تحسين دقة التدفق البصري، مما يساعد على تعزيز التناسق المكاني والزمني في الإطارات المتتالية.

على الرغم من فعاليةFRVSR ، فإنها لا تزال محدودة من حيث القدرة على إنتاج تفاصيل بصرية عالية الدقة وواقعية، خاصة في المشاهد المعقدة. هنا جاء دور خوارزمية **TecoGAN**، التي دمجت بين بنية FRVSR وتقنيات الشبكات التوليدية التنافسية (GAN)لتحسين جودة وتفاصيل الفيديو بشكل أكبر.

3.3 – TecoGAN (Temporally Coherent Generative Adversarial Network)

تعتمد خوارزمية[13] TecoGAN على استخدام نموذج شبكات توليدية تنافسية (GANs) مصمم خصيصًا لضمان جودة الإطارات من حيث التفاصيل المكانية والاتساق الزمني. تجمع TecoGAN بين الاستفادة من المعلومات المتسلسلة للإطارات السابقة، كما هو الحال في خوارزمية FRVSR، وإضافة شبكة تمييزية مكانية-زمانية (Spatio–Temporal Discriminator)لتحسين جودة وواقعية الإطارات الناتجة (Chu et al., 2018). علاوة على ذلك، تقدم الخوارزمية مفهومًا جديدًا يتمثل في تابع الخسارة Ping-Pong، الذي يهدف إلى تحسين التناسق الزمني طويل المدى، وتقليل التشوهات مثل الوميض والاهتزاز.

1.3.3– كيفية عمل الشبكات التوليدية التنافسية GANs

لتوضيح عمل TecoGan، من الضروري أن نفهم أولًا كيفية عمل GANs:

– GANs هي نوع من الشبكات العصبونية الاصطناعية التي تتكون من شبكتين فرعيتين رئيسيتين (Goodfellow et al. (2014:

1. **الشبكة التوليدية (Generator)**: وظيفتها توليد بيانات جديدة بناءً على بعض المدخلات، مثل توليد صور أو إطارات فيديو.
2. **الشبكة التمييزية (Discriminator)**: تعمل كمُصنِّف، حيث تقيم مدى "واقعية" البيانات التي تم توليدها من قبل الشبكة التوليدية، وتحاول التمييز بينها وبين البيانات الحقيقية.

– عملية التدريب في GANs تقوم على "التنافس" بين الشبكتين. الهدف من ذلك هو أن تتحسن الشبكة التوليدية تدريجيًا بحيث تصبح بياناتها الناتجة قريبة جدًا من البيانات الحقيقية لدرجة تجعل من الصعب على الشبكة التمييزية أن تميز بينهما. يتم حساب خسارة الخصم (Adversarial Loss) لكل من الشبكتين أثناء التدريب.

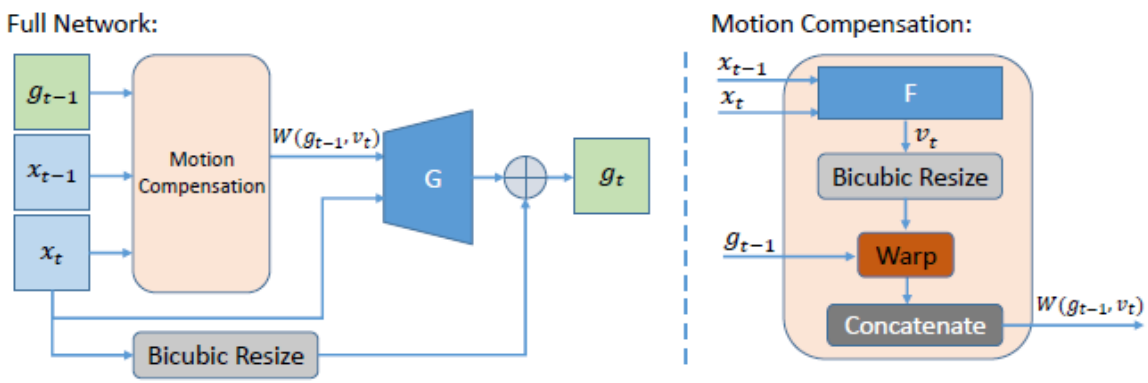
2.3.3- البنية العامة لخوارزمية TecoGan

تتكون البنية الأساسية لخوارزمية TecoGAN من ثلاثة مكونات رئيسية:

1. **الشبكة التوليدية التكرارية :(Recurrent Generator – SRNet)** مسؤولة عن تحسين الإطارات منخفضة الدقة وإنتاج إطارات عالية الدقة، مع الاستفادة من الإطارات الناتجة السابقة.
2. **شبكة تقدير الحركة :(Flow Estimation Network – FNet)** تُقدّر التدفق البصري بين الإطارات المتتابعة لتوفير مدخلات متناسقة زمنياً للشبكة التوليدية.
3. **الشبكة التمييزية المكانية–الزمانية :(Spatio–Temporal Discriminator – Ds,t)** تُقيّم جودة الإطارات الناتجة من حيث التفاصيل المكانية والتناسق الزمني.

سنقوم فيما يلي بشرح مكونات الخوارزمية بشكل مفَصل

1.2.3.3–الشبكة التوليدية التكرارية G :(Recurrent Generator – SRNet) تعتمد الشبكة التوليدية على هيكلية تكرارية. تأخذ الشبكة كمدخل الإطار الحالي منخفض الدقة X_t والإطار الناتج عالي الدقة السابق g_{t-1}، مع تعويض الحركة باستخدام خريطة الحركة v_tالمُقدرة بواسطة شبكة FNet



الشكل 5 : آلية عمل الشبكة التوليدية التكرارية (الجزء الأيسر)

آلية العمل :

– يتم استخدام الإطار الناتج السابق g_{t-1} بعد معالجته بتعويض الحركة. $W = (g_{t-1}, v_t)$ لتوفير سياق مكاني وزمني للإطار الحالي.

– يتم دمج المعلومات المستخلصة من الإطارات السابقة مع الإطار الحالي منخفض الدقة X_t لتوليد الإطار الجديد عالي الدقة g_t.

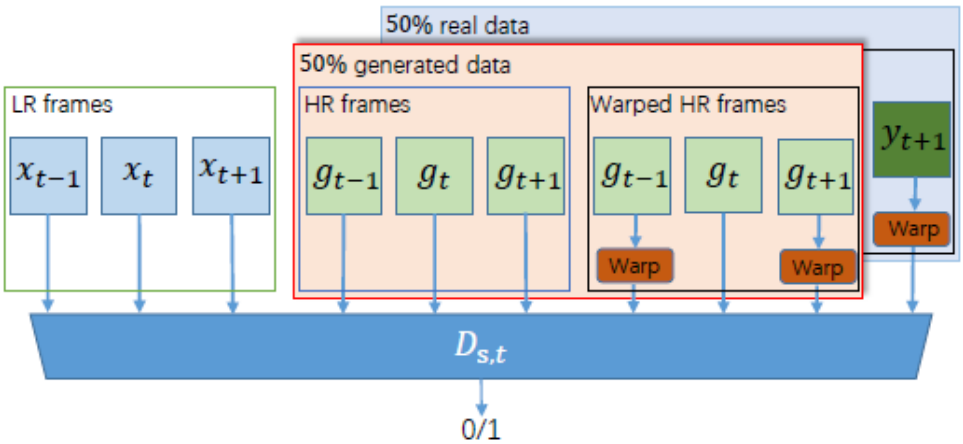
$$g_t = G\big(x_t, W(g_{t-1}, v_t)\big) + BicubicResize(X_t) \qquad \textbf{(21)}$$

يتم تدريب الشبكة على تعلم التفاصيل المتبقية(Residual Details) ، مما يساعد على تحسين استقرار التدريب والحد من الأخطاء الناتجة عن الإفراط في المواءمة Overfitting.

2.2.3.3– شبكة تقدير الحركة (Flow Estimation Network – FNet)

- يتم إدخال الإطارات منخفضة الدقة X_{t-1} و X_t إلى شبكة FNet، والتي تتكون من عدة طبقات تلافيفية (Convolutional Layers)مصممة لاستخراج الميزات المتعلقة بالحركة بين الإطارات ,مدعومة بطبقات تفعيل من نوع Leaky ReLU
- تقوم الشبكة بتحليل العلاقة المكانية–الزمانية بين الإطارين وتُنتج خريطة تدفق بصري v_t تمثل كيفية تحرك البيكسلات بين الإطارين.
- تُستخدم هذه الخريطة لتعويض الإطار السابق عالي الدقة g_{t-1} باستخدام تقنية تعويض الحركة (Motion Compensation)، مما يضمن أن الإطار الناتج الجديد g_t يتماشى مع الديناميكيات الزمنية للفيديو.

2.2.3.3– الشبكة التمييزية المكانية–الزمانية (Spatio–Temporal D Discriminator):



الشكل 6: المدخلات المستخدمة في المُميز المكاني–الزماني.

نلاحظ من الشكل 6 الشبكة التمييزية تستقبل مجموعتين من المدخلات:

1. مجموعة المدخلات الحقيقية (Real Inputs) :تتضمن الإطارات الحقيقية الطبيعية.
2. مجموعة المدخلات المولدة(Generated Inputs) : تتضمن الإطارات الناتجة عن الشبكة التوليديةG.

تحتوي كل مجموعة من المدخلات على:

- ثلاثة إطارات متوافقة منخفضة الدقة (LR) تم تكبيرها باستخدام (Bicubic Upsampling).
- ثلاثة إطارات متتالية عالية الدقة.(HR)
- ثلاثة إطارات مشوهة (Warped HR Frames) تم تعويضها باستخدام خريطة الحركة.

تلعب الشبكة التمييزية في TecoGAN دورًا أساسيًا في تقييم واقعية الإطارات الناتجة. على عكس الشبكات التمييزية التقليدية التي تقوم بتقييم الإطار بشكل فردي. تعتمد الشبكة التمييزية في TecoGAN على تحليل تسلسل الإطارات المتتابعة باستخدام تقنية PatchGAN. تقوم الشبكة بتقييم التفاصيل المكانية داخل كل إطار وكذلك التناسق الزمني بين الإطارات المتتالية.

فبدلًا من تحليل الصورة بالكامل دفعة واحدة، تقوم PatchGAN بتقسيم الصورة إلى أجزاء صغيرة (*patches*) وتحليل كل جزء على حدة. الهدف من هذه العملية هو تحسين التفاصيل المكانية على مستوى الأجزاء الصغيرة مثل الحواف والقوام وبالتالي تعزيز قدرة الشبكة على إنتاج إطارات ذات جودة عالية ومتناسقة زمنيًا. يتم خلال هذه العملية حساب جودة الإطار بناءً على مجموع جودة هذه الأجزاء الصغيرة.

3.3.3– توابع الخسارة المعتمدة في تدريب الخوارزمية TecoGAN

1. تابع الخسارة التربيعية المتوسطة:(MSE Loss)

تهدف هذه الخسارة إلى تقليل الفارق على مستوى البيكسل بين الإطارات الناتجة والإطارات المرجعية عالية الدقة:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left|\left|g_i - y_i\right|\right|^2 \qquad (22)$$

حيث :

y_i: الإطار الحقيقي عالي الدقة.

g_i: الإطار الناتج عن الشبكة التوليدية (عالي الدقة).

2.**تابع خسارة الخصائص:(Perceptual Loss)**

تُستخدم لتحسين جودة التفاصيل البصرية من خلال مقارنة الميزات المستخرجة من الإطارات المنتجة مع الإطارات المرجعية باستخدام شبكة VGG-19.

3. تابع خسارة الخصم:(Adversarial Loss)

تُستخدم لدفع الشبكة التوليدية إلى إنتاج إطارات أكثر واقعية:

$$L_{GAN} = \mathrm{E}_{y\sim p_{data}}\left[\log\left(D_{s,t}(y)\right)\right] + \mathrm{E}_{g\sim p_g}\left[\log\left(1 - D_{s,t}(g)\right)\right] \qquad \textbf{(23)}$$

حيث:

$D_{s,t}\ (y)$: احتمالية تصنيف الإطار الحقيقي على أنه حقيقي.

$D_{s,t}\ (g)$: احتمالية تصنيف الإطار الناتج على أنه حقيقي.

4. تابع الخسارة Ping-Pong :

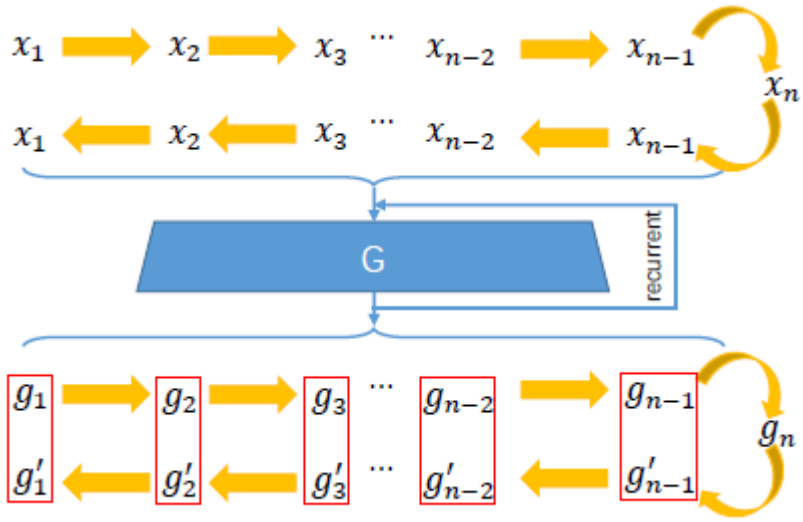
يعد تابع الخسارة Ping-Pongمن الابتكارات الجوهرية في خوارزمية TecoGAN، حيث تلعب دورًا حاسمًا في ضمان التناسق الزمني على المدى الطويل بين الإطارات المتتابعة. الهدف الرئيسي من تابع الخسارة هذا هو مقارنة الإطارات الناتجة عند تشغيل الفيديو في الاتجاه الأمامي وعند تشغيله بالعكس، للتأكد من أن الإطارات الناتجة تحافظ على الاتساق الزمني بغض النظر عن اتجاه عرضها. يساعد ذلك في تقليل التشوهات الزمنية مثل الوميض أو الاهتزاز، مما يجعل الفيديو يبدو سلسًا وطبيعيًا عند تشغيله .

$$L_{PP} = \sum_{i=1}^{N} \left|\left|g_i - g'_i\right|\right|^2 \qquad \textbf{(24)}$$

حيث :

g_i : الإطار الناتج عند تمرير التسلسل للأمام.

g'_i : الإطار الناتج عند تمرير التسلسل المعكوس.



الشكل 7 آلية عمل تابع خسارة **Ping-Pong**

4.1 آلية عمل خسارة Ping-Pong :

- **تمرير الإطارات للأمام**: يتم أولاً تمرير الإطارات منخفضة الدقة عبر الشبكة التوليدية لتحسينها، مما يولد الإطارات عالية الدقة **تمرير الإطارات بالإتجاه المعاكس** بعد تحسين الإطارات في الاتجاه الأمامي، يتم عكس ترتيب الإطارات الناتجة (بمعنى البدء من آخر إطار مُولد إلى أول إطار) ثم يُعاد تمريرها عبر الشبكة التوليدية مرة أخرى. في هذه الحالة، يتم تحسين الإطارات كما لو كان الفيديو يتم تشغيله في الاتجاه المعاكس، أي من النهاية إلى البداية.

مقارنة النتائج: تتم مقارنة الإطارات الناتجة في الاتجاهين (الأمام والخلف) للتأكد من أن التفاصيل المكانية والزمنية للإطارات الناتجة متطابقة في كلا الاتجاهين. يضمن هذا النهج أن الشبكة تحافظ على التناسق الزمني بين الإطارات بغض النظر عن اتجاه عرض الفيديو، مما يقلل من التشوهات مثل الوميض أو الاهتزاز الذي قد يحدث في الفيديو غير المتسق زمنيًا.

بفضل هذه الآلية، تضمن TecoGAN أن الإطارات المتتالية في الفيديو تظهر بشكل متناسق حتى عند تشغيل الفيديو في الاتجاه المعاكس، مما يعزز من سلاسة الحركة ويحد من التشوهات الزمنية التي قد تؤثر على جودة الفيديو. وبذلك، يتم الحفاظ على التناسق الزمني بين الإطارات في كلا الاتجاهين مما يحسن التجربة البصرية الإجمالية للفيديو.

5. الخسارة الإجمالية

تجمع TecoGAN بين هذه التوابع لتحقيق التوازن بين تحسين التفاصيل المكانية والتناسق الزمني:

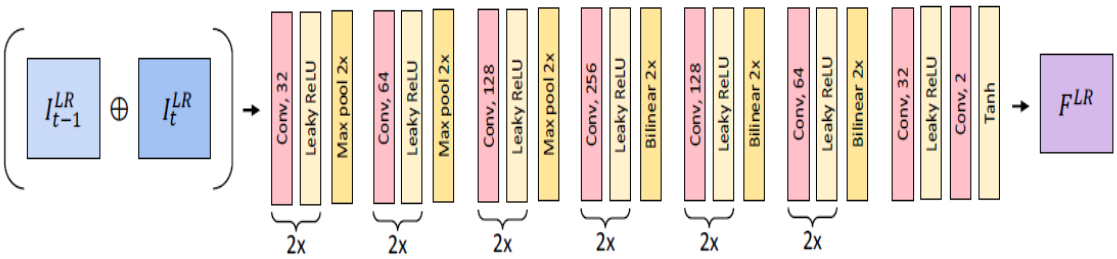
$$L_{total} = \lambda_{MSE} L_{MSE} + \lambda_P L_P + \lambda_{GAN} L_{GAN} + \lambda_{PP} L_{PP} \quad (24)$$

حيث :

λ_{MSE}, λ_p, λ_{GAN}, λ_{PP}, :عوامل الترجيح التي تُحدد مساهمة كل دالة خسارة

بعد شرح خوارزميتي FRVSR وTecoGAN، تظهر أهمية شبكة FNet التي تلعب دورًا محوريًا في تقدير الحركة بين الإطارات. دقة تقدير الحركة بواسطة FNet تؤثر بشكل مباشر على جودة الفيديو، حيث تعتمد الخوارزميتان على هذه الشبكة لضمان التناسق الزمني. لذا، تحسين FNet يعني إنتاج خرائط حركة أكثر دقة تؤدي إلى تقليل التشوهات الحركية مثل الوميض والاهتزاز، ما ينعكس بشكل إيجابي على جودة الإطارات المتتابعة .

4.3.3- بنية شبكة FNet



الشكل (8): بنية شبكة التدفق البصري FNet

شبكة FNet تتكون من عدة طبقات تلافيفية (Convolutional Layers) وطبقات تفعيل من نوع Leaky ReLU. الهدف من هذه الشبكة هو تقدير التدفق البصري بين الإطارات المتتابعة من الفيديو.

- **المدخلات** :تأخذ الشبكة إطارين منخفضي الدقة: الإطار الحالي $I_{LR,t}$والإطار السابق $I_{LR,t-1}$ وتعمل على تحليل الفرق بينهما.
- **الطبقات التلافيفية** :تتكون من عدة طبقات تلافيفية تركز على استخراج الميزات المتعلقة بالحركة بين الإطارات، حيث تعمل كل طبقة على تحسين فهم الشبكة لكيفية حركة البيكسلات.
- **طبقات** :Leaky ReLU تُستخدم لتمرير الإشارات الإيجابية مع تسريب بسيط للإشارات السلبية (0.2) لتحسين التعلم (المساعدة في تقليل الإفراط في المواءمة Overfitting).
- **الخرج** :يتمثل في خريطة التدفق البصري التي تمثل حركة البيكسلات بين الإطارين المدخلين.

5.3.3–تحويل FNet إلى شبكة عصبية ذات قيم عُقدية:(CVNNs)

الشبكات العصبونية التقليدية تعتمد على القيم الحقيقية، وهو ما قد يكون غير كافٍ عند معالجة بيانات معقدة مثل تدفق الحركة. وكما ذكرنا سابقاً خلال الدراسة النظرية يمكن للشبكات العصبونية ذات القيم العُقدية (CVNN) مثل الشبكة **FNet** تحسين تمثيل الحركة، وذلك بفضل قدرتها على التعامل مع المعلومات الطورية والاتجاهات المعقدة للحركة[15].

الفوائد الرئيسية لاستخدام القيم العُقدية تشمل :

1. **تمثيل دقيق للحركة:** يمكن للشبكات العُقدية تقديم تمثيل أفضل لمواقع الترددات العالية (التفاصيل الدقيقة ضمن الإطار) ، مما يؤدي إلى دقة أعلى في تقدير التدفق البصري[16].

2. **الاحتفاظ بالمعلومات الطورية:** الاحتفاظ بهذه المعلومات يعزز من دقة محاذاة الإطارات المتتالية[17].

3. **فهم أفضل للعلاقات الزمنية والمكانية:** يتيح استخدام الأعداد العُقدية للشبكة القدرة على فهم العلاقات المعقدة بين الإطارات المتتابعة، مما يعزز من دقة التنبؤات[15].

1.4.3.3–آلية التحويل :

1. المدخلات العُقدية:

يتم تحويل المدخلات (الإطارين منخفضي الدقة) إلى أعداد عُقدية .في هذه الحالة، يكون الجزء التخيلي للمدخل هو صفر، أي أن الإطارات المدخلة ستأخذ الشكل:

$$Z_{input} = I_{LR,t} + 0i \qquad (25)$$

2.الطبقات التلافيفية العُقدية:

الطبقات التلافيفية التقليدية تستبدل بطبقات تلافيفية عُقدية (ذات قيم عقدية) .حيث يتم استبدال المرشحات الحقيقية بمرشحات عُقدية:

$$Z_{output} = (W_r \,.\, X_r - W_i \,.\, X_i) + i(W_i . X_i + W_i \,.\, X_r) \qquad \textbf{(26)}$$

W_i , W_r : هما المعاملات (الأوزان) الحقيقية والتخيلية في المرشحات.

X_i , X_r : هما الجزء الحقيقي والتخيلي للدخل .

3. تابع التفعيل ذو القيم العُقدية:

تُستخدم توابع تفعيل عُقدية مثل Complex ReLU، حيث يتم تطبيق التفعيل على كلا الجزئين (الحقيقي والتخيلي).

4. الخرج:

بعد التحويل إلى قيم عُقدية، يكون الخرج أيضًا مركبًا. ومع ذلك، ولأن الخوارزميات تعتمد على قيم حقيقية، يجب أن يتم اختبار عدة سيناريوهات لضمان الحفاظ على المعلومات الطورية مع الاحتفاظ بقيم حقيقية كافية في الخرج.

1.4.خيارات معالجة الإخراج:

- بعد التحويل، يجب التعامل مع الخرج المركب لضمان توافقه مع الشبكة:

1. **تجاهل الجزء التخيلي:**
   - يمكن أن يتم تجاهل الجزء التخيلي في الخرج واستخدام الجزء الحقيقي فقط لتكوين خريطة التدفق البصري.
2. **دمج المعلومات الطورية:**
   - يمكن الاستفادة من المعلومات الطورية من الجزء التخيلي لتقديم خرائط أكثر دقة للحركة، مع ضمان أن تكون هذه الخرائط متوافقة مع باقي الشبكات الآخرى مثل SRNet.

الفصل الرابع

التطبيق العملي والنتائج

سيتم في هذا الفصل عرض خطوات التطبيق العملي للنموذج المقترح ونتائج التطبيق.

في هذا الفصل، نستعرض نتائج التطبيق العملي لتحويل الخوارزميات الأساسية إلى نسخ تعتمد على القيم العُقدية، حيث تم تطبيق هذا التحويل على خوارزمية SRCNN أولاً كدراسة اختبارية، ومن ثم على خوارزميتي FRVSR وTecoGAN لدراسة تأثير هذا التغيير على دقة النتائج والتناسق الزمني في الفيديوات. يظهر التحليل كيف أدى هذا التحويل إلى تحسينات ملموسة في جودة الإطارات الناتجة من حيث المعايير المختلفة دون الحاجة لزيادة كبيرة في عدد البارامترات، مما يدل على أن التحسن ناتج عن غنى المعلومات الإضافية التي توفرها القيم العُقدية.

1.4– نتائج خوارزمية SRCNN مع القيم العُقدية

1.1.4– إعداد التجربة

تم تدريب خوارزمية SRCNN بنسختها الأصلية وبنسختها المعدلة إلى شبكة عصبية ذات قيم عُقدية، وذلك باستخدام مرشحات متعددة لتحقيق تحسينات في الأداء. الهدف من هذا التحويل هو تقييم فعالية القيم العُقدية في تحسين جودة الصور الناتجة من خوارزميات تحسين دقة الصورة، مع الحفاظ على عدد بارامترات محدود لتجنب زيادة التعقيد الحسابي.

2.1.4– مجموعة الاختبار و مقاييس التقييم

تم اختيار مجموعة من مجموعات البيانات القياسية التي تُستخدم عادة في مهام تحسين دقة الصور الفردية، وتشمل تنوعًا في أنواع الصور مثل المشاهد الطبيعية والمشاهد الحضرية. تساعد هذه المجموعات في تقييم قدرة شبكة C-SRCNN (نموذج الشبكة SRCNN بقيم عُقدية) على التعميم ومعالجة محتويات الصور المختلفة، وتضم المجموعات التالية:

- **Set5** : تتكون من 5 أزواج من الصور منخفضة وعالية الدقة بدقة 256 256xبكسل.
- **Set14** : تتضمن 14 زوجًا من الصور منخفضة وعالية الدقة بدقة 512 512xبكسل.
- **BSD100** : تحتوي على 100 صورة عالية الدقة، وتستخدم عادةً لأغراض تحسين الدقة ومعالجة الصور.

- **Urban100** : تتألف من 100 صورة لمشاهد حضرية بدقة 512 512x بكسل، مما يجعلها ملائمة لتقييم التفاصيل في المشاهد المعقدة.

- **مقاييس التقييم**

لتقييم أداء C-SRCNN و SRCNNالتقليدية، اعتمدنا على المقاييس التالية: PSNR و SSIM

– تدريب الخوارزميتين SRCNN و CSRCNN

تم تدريب شبكة SRCNN وشبكة C-SRCNN باستخدام مجموعة التدريبT91 ومن أجل نفس العدد من الخطوات (Epoch)، التي تحتوي على 91 صورة، وتم إجراء التقييم على المجموعات القياسية المذكورة أعلاه

أظهرت النتائج أن شبكة CSRCNN تفوقت على الشبكة التقليدية SRCNN في جميع المقاييس تقريبًا. على سبيل المثال، حققت شبكة CSRCNN-915 تحسينًا في PSNR بمقدار 0.435 (تكبير X2)بالمقارنة مع-SRCNN 915.

Urban100	BSD100	Set14	Set5	Scale	النموذج
29.57502	32.83930	32.75063	35.8345	X2	SRCNN-915
-------	31.08669	31.32703	34.3566	X3	
26.89657	29.63494	29.55282	31.9265	X4	
29.85370	**33.04294**	**33.0407**	**36.3819**	X2	CSRCNN-915
-------	**31.20586**	**31.53459**	**34.6349**	X3	
26.99012	**29.70360**	**29.67355**	**32.0762**	X4	

جدول1 : PSNR

Urban100	BSD100	Set14	Set5	Scale	النموذج
0.908878	0.937895	0.91040	0.9512	X2	SRCNN-915
-------	0.903690	0.877280	0.9350	X3	
0.837133	0.871720	0.841073	0.9056	X4	
0.908821	**0.938649**	**0.91283**	**0.95369**	X2	CSRCNN-915
-------	**0.904986**	**0.879668**	**0.93757**	X3	
0.838977	**0.872626**	**0.842610**	**0.90751**	X4	

جدول 2: SSIM

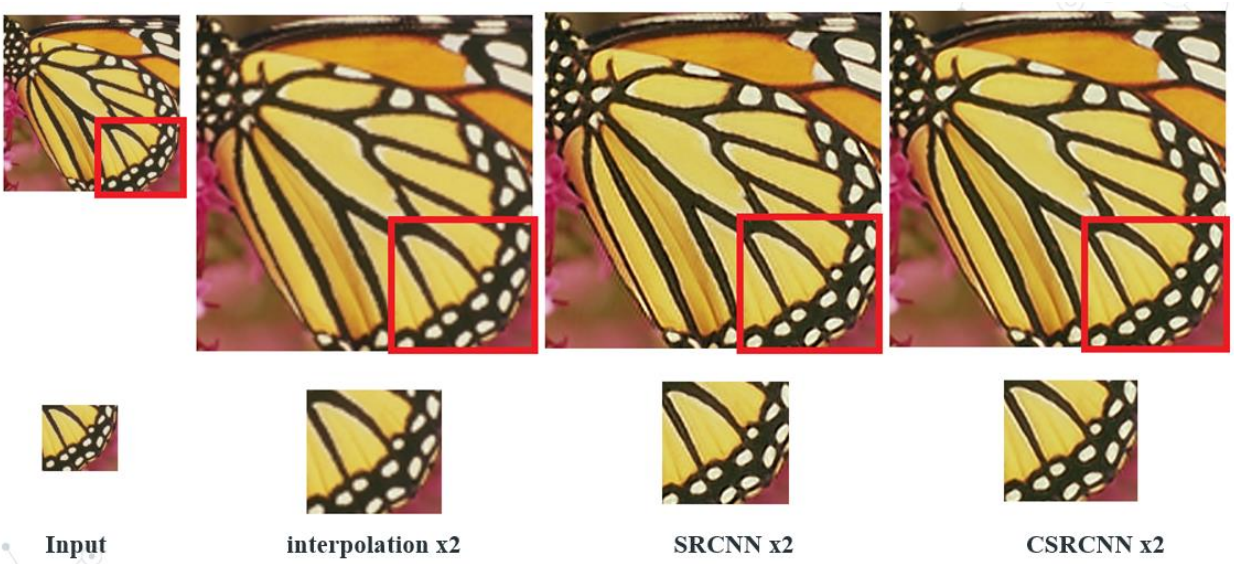
كما نلاحظ من الجدولين السابقين أن التحسن في قيم PSNR و SSIM هو أفضل من أجل زيادة الدقة أو التكبير بنسبة 2 (X2) وهذا منطقي نظراً لزيادة صعوبة استرجاع تفاصيل أدق مع زيادة الدقة بنسبة أكبر أي X3 و X4.

تُظهر النتائج في الجدول (3) أن النسخة المعدلة **CSRCNN** قد حققت أداءً محسّنًا في جميع المقاييس تقريبًا مقارنةً بالنسخ الأصلية، مما يشير إلى أن استخدام القيم العُقدية في الشبكة أدى إلى تعزيز جودة الصور المولدة. هذا التحسن لم يكن بسبب زيادة عدد البارامترات، بل كان نتيجة لزيادة غنى المعلومات المستخلصة من القيم العُقدية.

كما يظهر الشكل (9) مقارنة بصرية بين نتائج الخوارزميات الثلاث: الإستيفاء التكعيبي و SRCNN و CSRCNN حيث يبدو واضحاً تفوق الخوارزميتن SRCNN و CSRCNN على خوارزمية الاستيفاء التكعيبي ولكن يظهر أيضا صعوبة المقارنة البصرية بين الخوارزميتين SRCNN و CSRCNN وهذا يظهر اهمية استخدام مقاييس التشابه الكمية PSNR و SSIM معاً.

Urban100	BSD100	Set14	Set5	Scale	النموذج
29.57502	32.83930	32.75063	35.8345	X2	SRCNN-915 Parameters : 20.099
-------	31.08669	31.32703	34.3566	X3	
26.89657	29.63494	29.55282	31.9265	X4	
29.7602	32.9435	33.0301	36.3159	X2	SRCNN-935 Parameters : 36.483
-------	31.1034	31.3659	34.4074	X3	
26.8977	29.6133	29.5404	31.9210	X4	
29.6711	32.9076	32.9502	36.0525	X2	SRCNN-955 Parameters : 69.251
-------	31.0606	31.2873	34.3292	X3	
26.8787	29.6126	29.5225	32.9078	X4	
29.85370	**33.04294**	**33.0407**	**36.3819**	X2	CSRCNN-915 Parameters : 40.198
-------	**31.20586**	**31.53459**	**34.6349**	X3	
26.99012	**29.70360**	**29.67355**	**32.0762**	X4	

جدول 3: PSNR , مع تغيير المرشحات و تبيان عدد البارامترات لكل شبكة



الشكل (9): المقارنة البصرية لنتائج الخوارزميتين SRCNN و CSRCNN

2.4– إعداد التجربة لخوارزمياتFRVSR و Tecogan

1.2.4– مجموعة التدريب

تم تنفيذ عملية التدريب على مجموعة REDS، والتي تتضمن مقاطع فيديو عالية الجودة. شملت مجموعة التدريب 30 مقطع فيديو، بينما تم تخصيص 10 مقاطع أخرى لمجموعة التحقق(validation) ، حيث يتألف كل مقطع من 100 إطار(بدقة 1280×720 بكسل).

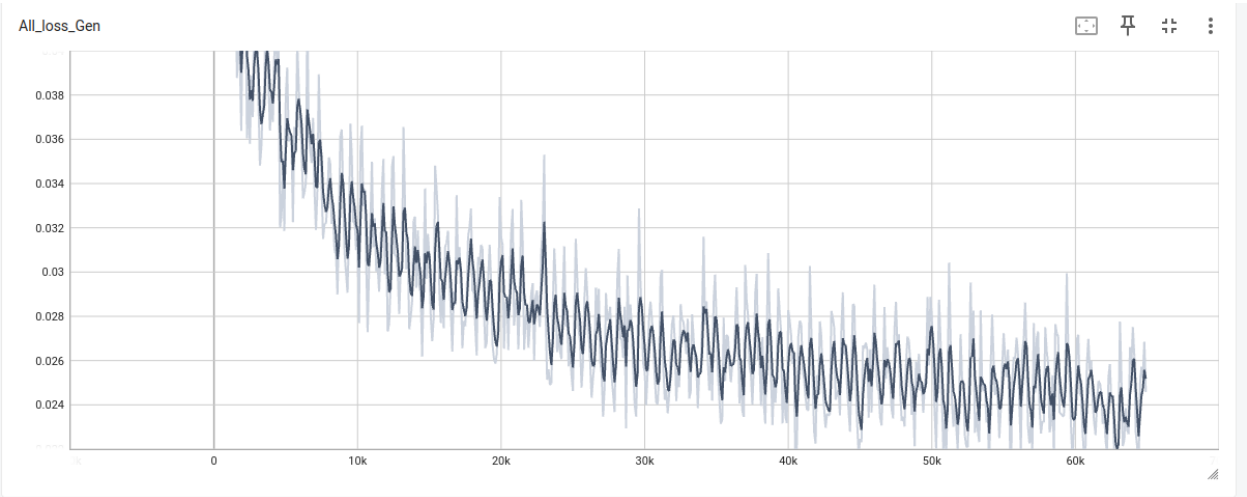
2.2.4– تدريب الخوارزميتين FRVSR و TecoGAN

أظهر منحنى التدريب لكلا النموذجين أن الأداء ما زال قابلاً للتحسين بعد هذه الخطوات، مما يشير إلى إمكانية الحصول على نتائج أفضل إذا تم تمديد عملية التدريب.

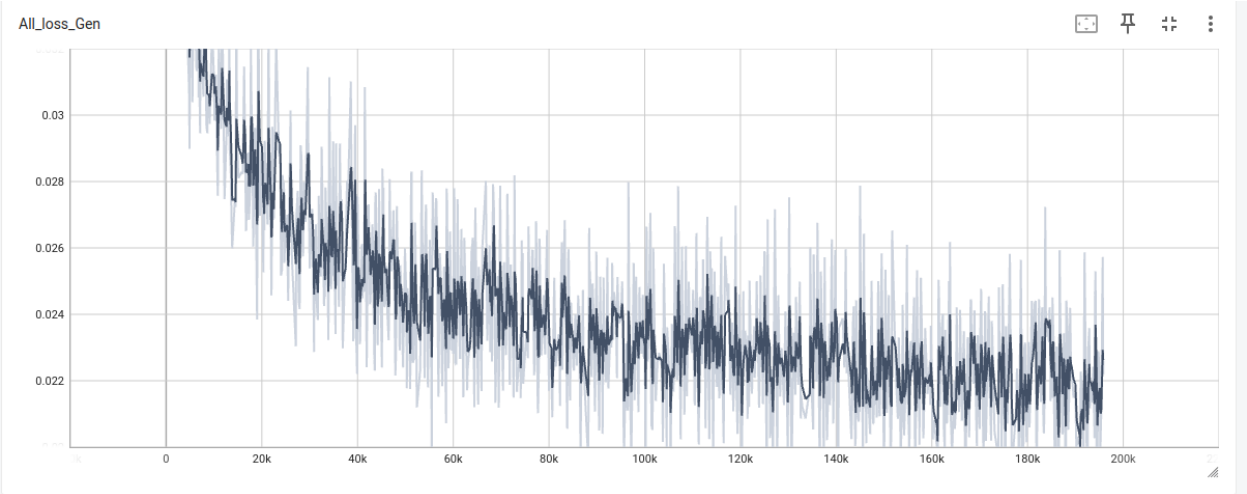
تم تدريب خوارزمية FRVSR إلى 190,000 خطوة تدريبية (steps) ، معدل التعلم.0.0005 :(Learning Rate)

حجم الدُفعة.4 :(Batch Size) , طول التسلسل 10 :(RNN_N) إطارات., عدد الكتل المتبقية (Residual Blocks): 10.

تم التدريب على حاسب نظام Linux , .NVIDIA GTX 1060 (GPU) , استغرق زمن التدريب 157ساعة للشبكات العصبونية ذات القيم الحقيقية كما تم تدريب الشبكات العصبونية ذات القيم العُقدية على حاسب نظام windows10 , Geforce RTX 2080 (GPU) , استغرق زمن التدريب 188 ساعة.



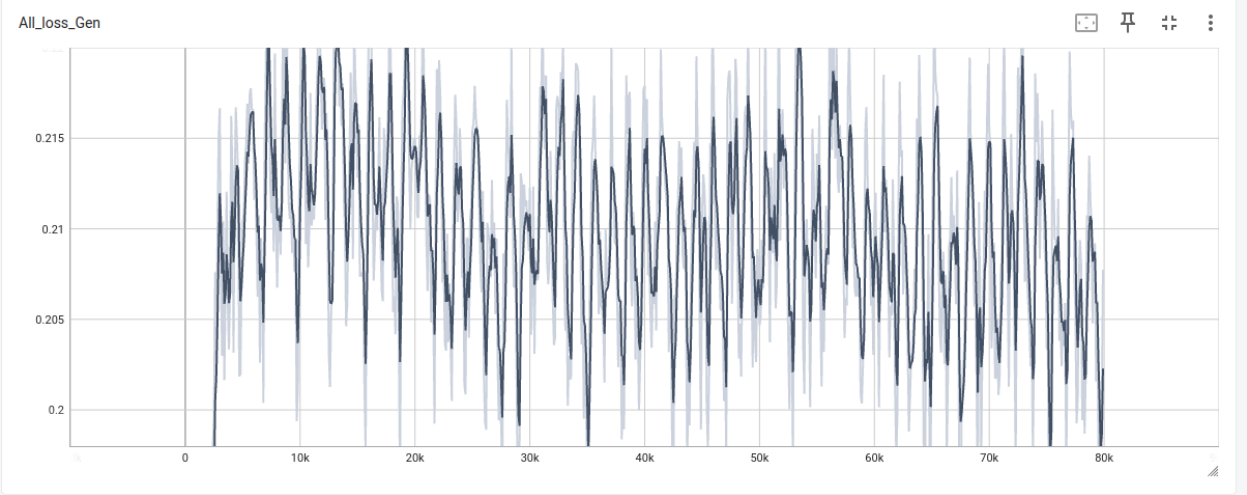
الشكل10 : منحني تدريب FRVSR



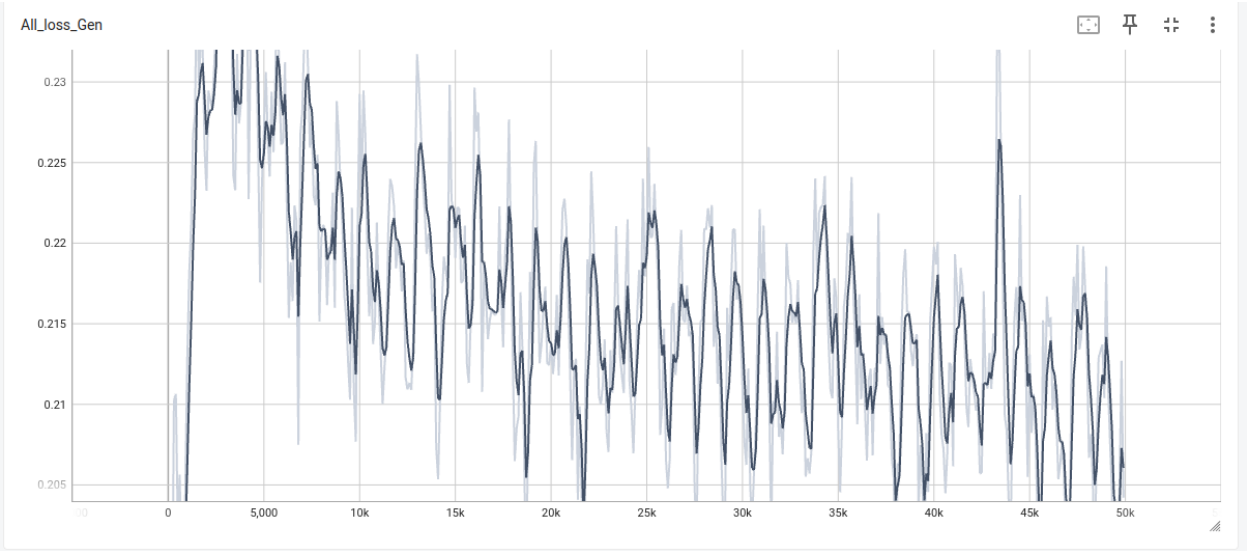
الشكل 11: منحني تدريب Complex-FRVSR

بينما تم تدريب خوارزمية TecoGAN إلى 500,000 خطوة تدريبية (steps) ، معدل التعلم :(Learning Rate) 0.0005. و حجم الدُفعة.4 :(Batch Size) , طول التسلسل 10 :(RNN_N) إطارات., عدد الكتل المتبقية (Residual Blocks): 16.

تم التدريب على حاسب نظام windows10 , Geforce RTX 2080 (GPU) , استغرق زمن التدريب الشبكات ذات القيم الحقيقية 149 ساعة بينما استغرق تدريب الشبكات العصبونية العقدية 236 ساعة .



الشكل 12: منحني تدريب Tecogan



الشكل 13: منحني التدريب Complex-Tecogan

2.2.4- المقاييس المستخدمة لتقييم FRVSR وTecoGAN

لتقييم أداء خوارزميتي FRVSR وTecoGAN بعد تحويل FNet إلى شبكة ذات قيم عُقدية، تم استخدام المقاييس التالية:

- PSNR : يقيس نسبة الإشارة إلى الضجيج، ويعكس جودة التفاصيل في الفيديوات الناتجة.
- LPIPs : يعبر عن الفروقات الإدراكية بين الإطارات، حيث تشير القيم الأقل إلى تقارب الإدراك البشري بين الإطارات المولدة والإطارات الأصلية.
- tof : يقيس التناسق الزمني بين الإطارات المتتالية، مما يساعد في تقييم مدى سلاسة الحركة.
- tLp100 : يقيم التناسق الإدراكي على مدى الإطارات المتتابعة لضمان خلو الفيديو من التشوهات الزمنية.

3.2.4- نتائج FRVSR وComplex-FRVSR

تم اختبار خوارزميتي FRVSR وTecoGAN على مجموعة بيانات Vid4 تُستخدم على نطاق واسع في اختبار خوارزميات تحسين دقة الفيديو (Video Super-Resolution). تحتوي هذه المجموعة على أربع تسلسلات فيديو تتميز بتنوع المشاهد وتحديات في التفاصيل الحركية والزمني:

- **Walk** : بدقة 740×480 وبعدد إطارات يبلغ 47 إطارًا.
- **Foliage** : بدقة 740×480 وبعدد إطارات يبلغ 49 إطارًا.
- **City** : بدقة 704×576 وبعدد إطارات يبلغ 34 إطارًا.

- **Calendar** : بدقة 720×576 وبعدد إطارات يبلغ 41 إطارًا.

يوضح الجدول التالي نتائج خوارزمية FRVSR والنسخة المعدلة Complex–FRVSR بعد التحويل إلى قيم عُقدية:

النموذج	معيار	vid4	مقدار التحسن %
FRVSR	PSNR	25.8709	
	LPIPs	0.298325	
	tof	0.248275	1.78
	tLp100	1.2028	
Complex-FRVSR	PSNR(db)	25.969	0.38
	LPIPs	0.28575	-4.21
	tof	0.2527	
	tLp100	1.163325	-3.28

الجدول (4) : نتائج خوارزمية FRVSRوالنسخة المعدلة Complex–FRVSR على مجموعة الاختبار vid4

كما تم الاختبار على مجموعة البيانات Tos3 حيث تحتوي على مشاهد ذات حركات كاميرا سلسة وأخرى معقدة، تتألف هذه المجموعة من ثلاث تسلسلات فيديو تتضمن تحديات بصرية مثل الإضاءة المتنوعة والحركة.

- Face : بدقة 1280×534 وبعدد إطارات يبلغ 166 إطارًا.
- Room : دقة 1280×534 وبعدد إطارات يبلغ 150 إطارًا.
- Bridge : دقة 1280×534 وبعدد إطارات يبلغ 233 إطارًا.

يوضح الجدول التالي نتائج خوارزمية FRVSR والنسخة المعدلة Complex–FRVSR بعد التحويل إلى قيم عُقدية:

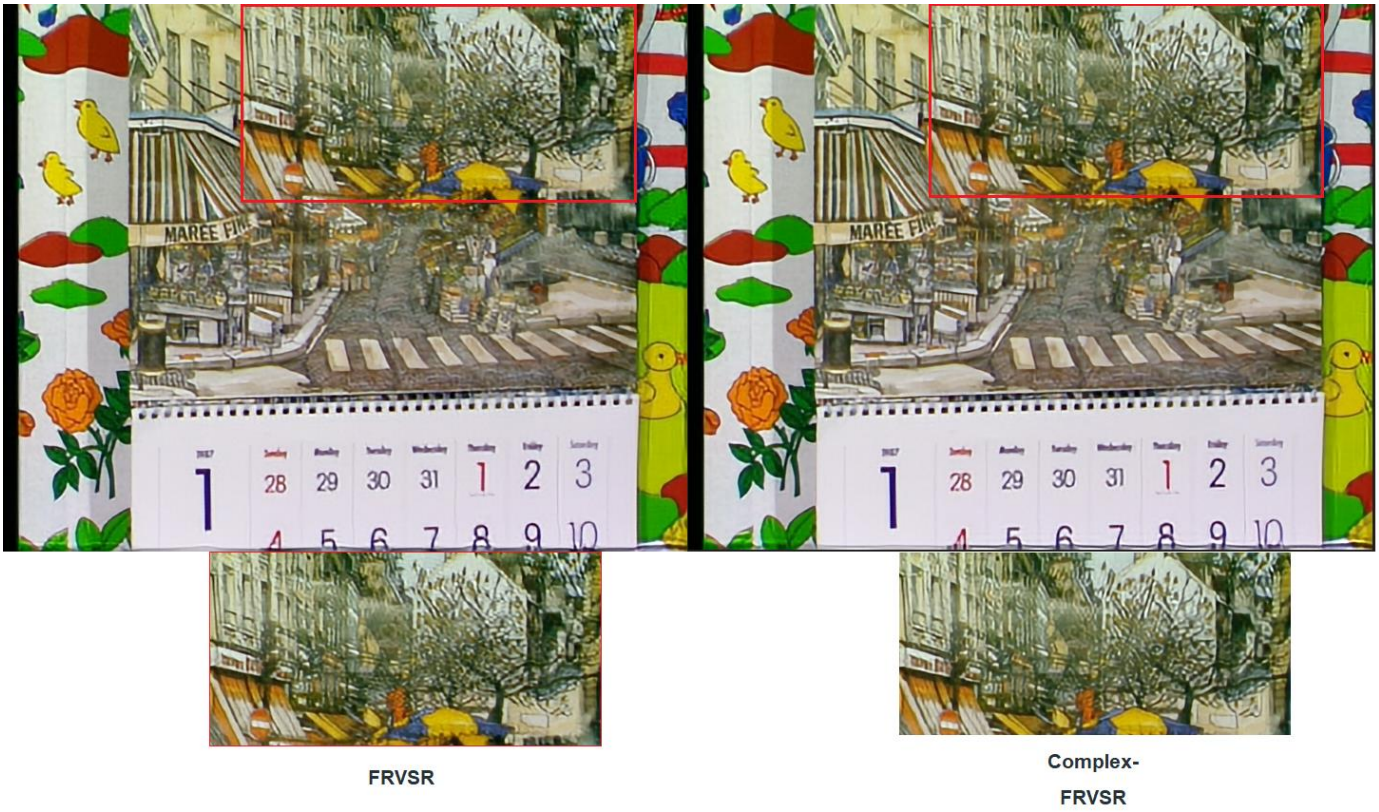
النموذج	معيار	tos3	مقدار التحسن %
FRVSR	PSNR	25.8650	
	LPIPs	0.365	
	tof	0.6703	
	tLp100	1.5707	
Complex-FRVSR	PSNR(db)	31.6719	22.45
	LPIPs	0.1835	-49.72
	tof	0.173866	-74.06
	tLp100	0.448766	-71.43

الجدول (5) : نتائج خوارزمية FRVSR والنسخة المعدلة Complex–FRVSR على مجموعة الاختبار tos3

تشير النتائج إلى أن Complex–FRVSR حققت بشكل عام تحسينات في جميع المقاييس. خاصة في tos3 ,حيث تحسن PSNR بشكل كبير، وأصبحت الجودة البصرية أكثر وضوحًا، وتحسن الاتساق الزمني والتدفق البصري بشكل كبير. وبالنظر إلى منحنيات التدريب المعروضة سابقاً في الشكل (11)، والتي بينت إمكانية التدريب لعدد أكبر بكثير من الخطوات حيث لاتزال الخوارزمية في طور التقارب، نرى ضرورة تدريب الخوارزمية لعدد أكبر بكثير من الخطوات والذي سيؤدي باعتقادنا لتحسن كبير وملموس في النتائج بشكلها الكمي (مقاييس التشابه) والنوعي (المقارنة البصرية).

input / LR

180*144



FRVSR **Complex-FRVSR**

الشكل (14): المقارنة البصرية لنتائج الخوارزميتين FRVSR و Complex-FRVSR

4.2.4- نتائج Tecogan وComplex-TecoGAN

تم توضيح نتائج خوارزمية TecoGAN والنسخة المعدلة Complex-TecoGAN في الجدولين التاليين :

النموذج	معيار	vid4	مقدار التحسن %
TecoGan	PSNR	24.8445	
	LPIPs	0.1795	
	tof	0.2302	
	tLp100	0.64195	4.50
Complex-TecoGan	PSNR(db)	25.0759	0.93
	LPIPs	0.176125	-1.88
	tof	0.228725	-0.64
	tLp100	0.670825	

جدول(6) نتائج خوارزمية TecoGAN والنسخة المعدلة Complex-TecoGAN على مجموعة الاختبار vid4

النموذج	معيار	tos3	مقدار التحسن %
TecoGan	PSNR	29.3735667	
	LPIPs	0.18493333	
	tof	0.17523333	
	tLp100	0.308	
Complex-TecoGan	PSNR(db)	30.4241333	3.58
	LPIPs	0.1592	-13.91
	tof	0.16736667	-4.49
	tLp100	0.2838333	-7.85

جدول(7) نتائج خوارزمية TecoGAN والنسخة المعدلة Complex–TecoGAN على مجموعة الاختبار tos3

تشير النتائج إلى أن Complex–TecoGAN حققت تحسينًا ملحوظًا، خصوصًا في مقاييس التناسق الزمني، مما يعزز دقة التفاصيل والتناسق عبر الإطارات. كما تشير النتائج أيضاً لضرورة التدريب لعدد أكبر من الخطوات للحصول على تحسن كبير وملموس في النتائج بشكلها الكمي (مقاييس التشابه) والنوعي (المقارنة البصرية).

input / LR
320*134

الشكل (15): المقارنة البصرية لنتائج الخوارزميتين TecoGan و Complex-TecoGan

5.2.4– التحليل العام

- **فعالية القيم العُقدية** :تشير النتائج لكل من CSRCNN و Complex–FRVSR و Complex–TecoGAN إلى أن استخدام القيم العُقدية يضيف غنىً معلوماتيًّا يسهم في تحسين جودة التفاصيل والتناسق، دون الحاجة إلى زيادة كبيرة في عدد البارامترات.
- **التحديات والقيود** :أدى استخدام القيم العُقدية إلى تعقيد الحسابات، إلا أن التحسينات التي تم تحقيقها تثبت أن هذا التعديل يستحق الاستخدام في التطبيقات التي تتطلب جودة عالية.

تم في هذا الفصل عرض نتائج التحسينات التي أدخلت على خوارزميات SRCNN و FRVSR و TecoGAN عبر استخدام القيم العُقدية، وتبين أن هذه التحسينات أسهمت في تحسين جودة الفيديوات والصور الناتجة من حيث الدقة والتناسق الزمني والمكاني.

الفصل الخامس

الخاتمة والآفاق المستقبلية

1.5-الخاتمة:

في هذا البحث، تم استكشاف تأثير استخدام القيم العُقدية في شبكات التعلم العميق لتحسين دقة الصور والفيديوات، وذلك من خلال تعديل عدد من الخوارزميات التقليدية مثل SRCNN و FRVSR و TecoGANلتصبح شبكات تعتمد على القيم العُقدية. تناول البحث تحليلًا تفصيليًا لأداء هذه الشبكات بعد التحويل إلى قيم عُقدية، وتمت المقارنة مع النسخ التقليدية لكل خوارزمية باستخدام معايير شائعة مثل PSNRو SSIMوLPIPs، بالإضافة إلى معايير التناسق الزمني مثل tof و tLp100 .

أظهرت النتائج أن الشبكات العصبونية ذات القيم العُقدية قدمت أداءً محسّنًا مقارنةً بالشبكات التقليدية، مما يعكس فعالية القيم العُقدية في تعزيز جودة الصور والفيديوات الناتجة من حيث الدقة والتفاصيل والتناسق الزمني. يعد هذا التحسن نتيجة لغنى المعلومات الإضافية التي توفرها القيم العُقدية، مما يعزز قدرة النموذج على التقاط التفاصيل الدقيقة ومعالجة التغيرات الزمنية في الفيديوات. وبذلك، يعتبر هذا البحث خطوة مهمة نحو تحسين خوارزميات تحسين دقة الصور والفيديو باستخدام تقنيات التعلم العميق.

2.5-الآفاق المستقبلية:

بالنظر إلى النتائج الإيجابية التي حققها هذا البحث، يمكن توجيه الجهود المستقبلية نحو تطوير النموذج ليعمل في الزمن الحقيقي، وهو ما يستلزم تحسين كفاءة الشبكات العصبونية وتقليل الزمن الحسابي مع الحفاظ على مستويات الأداء العالية من حيث الجودة والدقة. كما يمكن العمل على تعديل النموذج ليصبح متوافقاً مع بيئات الحوسبة المصغرة، مثل حواسيبJetson Orin ، من خلال تحسين استهلاك الموارد الحاسوبية وضمان الاستفادة المثلى من قدرات المعالجة المتاحة على هذه الأجهزة.

المراجع

1. Keys, R. G. (1981). Cubic convolution interpolation for digital image processing *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6), 1153-1160
2. .Duchon, C. E. (1979). Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology and Climatology*, 18(8), 1016-1022.
3. Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4), 259-268.
4. Freeman, W. T., Jones, T. R., & Pasztor, E. C. (2002). Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2), 56-65.
5. Yang, J., Wright, J., Huang, T., & Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11), 2861-2873.
6. Chang, H., Yeung, D. Y., & Xiong, Y. (2004). Super-resolution through neighbor embedding. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004, Vol. 1, 275-282.
7. Liu, C., & Sun, D. (2011). A Bayesian approach to adaptive video super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 209-216.
8. Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K. (2016). Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging*, 2(2), 109-122.
9. Caballero, J., Ledig, C., Aitken, A. P., Acosta, A., Totz, J., Wang, Z., & Shi, W. (2017). Real-time video super-resolution with spatio-temporal networks and motion compensation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4778-4787.
10. Wang, X., Chan, K. C., Yu, K., Dong, C., & Change Loy, C. (2019). EDVR: Video restoration with enhanced deformable convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Workshops, 0-0.
11. Sajjadi, M. S. M., Vemulapalli, R., & Brown, M. (2018). Frame-recurrent video super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6626-6634.
12. Haris, M., Shakhnarovich, G., & Ukita, N. (2019). Recurrent back-projection network for video super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3897-3906.

13. Chu, X., Xie, Z., Wu, J., Zheng, W., & Ma, H. (2021). TecoGAN: Temporally coherent GAN for video super-resolution. *IEEE Transactions on Image Processing*, 30, 3885-3898.
14. Zeng, R., Huang, W., Fu, J., Fu, H., Chao, H., & Luo, J. (2020). Learning spatio-temporal transformer network for motion-aware video super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1316-1325.
15. Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., & Pal, C. J. (2018). Deep complex networks. *International Conference on Learning Representations (ICLR)*, 2018.
16. Virtue, P., & Lustig, M. (2016). Complex-valued convolutional neural networks for MRI reconstruction. *IEEE Conference on Image Processing (ICIP)*, 0-0.
17. Hirose, A., & Yoshida, K. (2012). Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Transactions on Neural Networks and Learning Systems*, 23(4), 541-551.
18. Zeng, R., Huang, W., Fu, J., Fu, H., Chao, H., & Luo, J. (2020). Learning spatio-temporal transformer network for motion-aware video super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1316-1325.
19. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *International Conference on Computer Vision (ICCV)*, 10012-10022.
20. Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5), 529-541
21. C. Dong, C. C. Loy, K. He and X. Tang, "Image super-resolution using deep convolutional networks", IEEE Transactions on pattern analysis and machine intelligence, vol. 38, pp. 295-307, 2016.
22. J. Kim, J. K. Lee and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks", CVPR, 2016.
23. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems (NeurIPS)*, 27, 2672-2680.
24. M. Schwertfeger, A. N. Gholami, and M. Lustig, "Theory and applications of complex-valued convolutional networks," *Conference on Learning Representations (ICLR)*, 2021.
25. A. Gholami, M. Schwertfeger, and Y. Bengio, "A survey on complex-valued neural networks," *Neural Information Processing Systems (NeurIPS)*, 2019.

Abstract

Super-Resolution (SR) is a significant technique in image and video processing, aiming to enhance the resolution of low-quality images, transforming them into high-resolution ones, thereby improving detail and clarity. Most models for image and video resolution enhancement, especially those based on neural networks, rely on real values for representing input images and subsequent processing stages. This approach often overlooks phase information, which could be essential for achieving more accurate image representation.

This research aims to explore the potential of complex-valued neural networks, enabling the incorporation of phase information alongside magnitude, thus enhancing the model's ability to represent fine details more effectively. We modified existing algorithms based on real-valued neural networks, such as SRCNN, FRVSR, and TecoGAN, to operate using complex values. The performance of these models was evaluated using diverse benchmark datasets encompassing natural and urban scenes, aiming to assess the quality improvement of the generated images and videos. The evaluation employed commonly used metrics, such as PSNR and SSIM, for accuracy, as well as temporal metrics like tof and tLp100 to ensure temporal consistency across consecutive frames.

The results demonstrated that complex-valued neural networks achieved significant improvements compared to their real-valued counterparts, delivering superior performance in image resolution enhancement and temporal consistency in videos without incurring substantial increases in parameter count. This indicates that leveraging phase information through complex values can enhance the quality of results and improve the efficiency of super-resolution algorithms in image and video processing.

Keywords: Super-Resolution, Deep Neural Networks, Complex Values, Phase Information, SRCNN, FRVSR, TecoGAN, PSNR, SSIM.

Syrian Arab Republic
Higher Institute of Applied Science and Technology
Informatics Department

A thesis submitted for the degree of
Master in Big Data Systems

A model for video super-resolution using complex-valued neural network

Submitted by
Hilda Zaher

Supervised by
Dr. Assef Gafar **Dr. Bassem alsahwa**

2024