



**HỌC VIỆN
CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

Posts and Telecommunications Institute of Technology

BÁO CÁO THỰC TẬP CƠ SỞ

ĐỀ TÀI: LÀM RÕ HÌNH ẢNH SÂU

Sinh viên thực hiện: Nguyễn Thị Lan

Giáo viên hướng dẫn: ThS Đinh Xuân Trường

Mã sinh viên: B21DCCN818 - Lớp: D21CQCN09-B

Khoa CNTT 1 - PTIT

23/04/2024

Nội dung

1. Giới thiệu

2. Cơ sở lý thuyết

2.1 Phát biểu vấn đề

2.2 Đánh giá chất lượng ảnh

3. Làm Rõ Ảnh Không Mù

4. Làm Rõ Ảnh Mù

5. Các hàm mất mát

6. Các Bộ Dữ liệu Thử Nghiệm cho Việc Làm Rõ Ảnh

6.1 Bộ dữ liệu làm rõ ảnh

6.2 Bộ dữ liệu làm rõ video

6.3 Bộ dữ liệu cụ thể cho từng lĩnh vực

7. Đánh giá hiệu suất

8. Làm Rõ Ảnh Theo Miền Cụ Thể

Giới thiệu

Giới thiệu chung

Làm mờ hình ảnh là một nhiệm vụ kinh điển trong thị giác máy tính cấp thấp, đã thu hút được sự chú ý từ cộng đồng xử lý hình ảnh và thị giác máy tính. Mục tiêu của việc làm mờ hình ảnh là khôi phục hình ảnh sắc nét từ hình ảnh đầu vào bị mờ, trong đó hiện tượng mờ có thể do nhiều yếu tố khác nhau gây ra như thiếu tiêu điểm, rung máy hoặc chuyển động mục tiêu nhanh



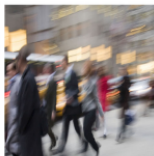
(a) Camera shake blur



(b) Out-of-focus blur



(c) Moving object blur



(d) Mixed blur

Hình: Ảnh mờ

Cơ sở lý thuyết

Phát biểu vấn đề

Mờ ảnh có thể phát sinh từ nhiều nguyên nhân khác nhau trong quá trình chụp ảnh, bao gồm rung máy ảnh, chuyển động trong cảnh, và mất nét. Mục tiêu chính của việc làm rõ ảnh là khôi phục lại ảnh sắc nét từ ảnh bị mờ.

- **Mờ Chuyển Động:** Một nguyên nhân phổ biến của mờ ảnh là do chuyển động trong quá trình chụp ảnh, khi thời gian mở ống kính dài hơn dẫn đến mờ chuyển động. Các phương pháp xử lý này thường dựa trên mô hình tích chập và nhiễu Gaussian.
- **Mờ Lấy Nét:** Một nguyên nhân khác là mất nét do khoảng cách giữa cảnh và mặt phẳng tiêu điểm của máy ảnh. Các phương pháp xử lý mờ lấy nét thường sử dụng hàm phân tán điểm (PSF) để mô hình hóa quá trình mất nét.
- **Mờ Gaussian và Mờ Kết Hợp:** Mờ Gaussian là một mô hình phổ biến được sử dụng để mô tả mờ trong xử lý ảnh. Trong một số trường hợp, mờ có thể là kết hợp của nhiều yếu tố như rung máy ảnh, chuyển động đối tượng và mất nét.

Đánh Giá Chất Lượng Ảnh

Các phương pháp đánh giá chất lượng ảnh (IQA) có thể được phân loại thành các chỉ số chủ quan và khách quan. Các phương pháp chủ quan dựa trên đánh giá của con người, có thể không yêu cầu một ảnh tham chiếu. Một chỉ số đại diện là Điểm Ý Kiến Trung Bình (MOS), trong đó người dùng đánh giá chất lượng của ảnh trên một thang điểm từ 1-5. Bởi MOS phụ thuộc vào mẫu dân số, các phương pháp thường tính đến các thống kê của các điểm ý kiến. Đối với việc làm rõ ảnh, hầu hết các phương pháp hiện có được đánh giá dựa trên các chỉ số đánh giá khách quan, có thể được chia thành hai loại: chỉ số IQA có tham chiếu đầy đủ và không có tham chiếu.

Làm Rõ Ảnh Không Mùa

Làm rõ ảnh không mờ

Trong

trường hợp hạt nhân mờ được cung cấp, vấn đề được gọi là làm rõ ảnh không mờ. Tuy nhiên, ngay cả khi có sẵn hạt nhân mờ, nhiệm vụ vẫn gặp khó khăn do nhiễu cảm biến và sự mất thông tin tần số cao. Các phương pháp không mờ:

- Giải Phóng Không Mờ Kèm Làm Sạch Nhiễu
- Học Mô-đun Làm Sạch Nhiễu cho Làm Rõ Ảnh Không Mờ

Method	Category	Blur type	Dataset	Architecture	Key idea
DCNN Xu et al. (2014)		Gaussian, disk			First work to combine traditional optimization-based schemes and neural networks
BRCNN Zhang et al. (2017)		Gaussian, motion			Learn a set of CNN denoisers, use as a modular part of model-based optimization methods to tackle inverse problems
FCNN Zhang et al. (2017)		Motion			Adaptively learn image priors to preserve image details and structure with a robust L1 loss
FDN Kruse et al. (2017)	Uniform	Motion	Convolution	CNN	learns a CNN-based prior with an FFT-based deconvolution scheme
GLRA Ren et al. (2018)		Gaussian, disk, motion			Use generalized low-rank approximations of blur kernels to initialize the CNN parameters
DUBLID Li et al. (2019)		Motion			Recast a generalized TV-regularized algorithm into a deep network for blind image deblurring
RGDN Gong et al. (2020)		Motion			Incorporate deep neural networks into a fully parameterized gradient descent scheme
DWDN Dong et al. (2020)		Motion, Gaussian			Apply explicit deconvolution in feature space by integrating a classical Wiener deconvolution framework
USRNet Zhang et al. (2020)		Motion, Gaussian			End-to-end training of an unfolding network that integrates advantages of model-based and learning-based methods

Hình: Khái quát về các phương pháp làm mờ ảnh đơn hình sâu không mờ đơn

Làm Rõ Ảnh Mù

Đầu vào của mạng và tổng hợp khung hình

Cụ thể, mạng làm rõ ảnh đơn nhận một ảnh mờ duy nhất và tạo ra kết quả làm rõ tương ứng, trong khi các phương pháp làm rõ video sử dụng nhiều khung hình và tổng hợp thông tin từ các khung hình này. Các thuật toán tổng hợp ở mức độ ảnh xếp nhiều khung hình và ước lượng kết quả cho khung hình trung tâm, trong khi ở mức độ đặc trưng, trích xuất đặc trưng từ các khung hình và hợp nhất chúng để dự đoán kết quả làm rõ.

Các lớp và khối cơ bản được sử dụng trong mạng làm rõ ảnh.

Các lớp và khối này bao gồm:

Lớp tích chập

Lớp đệ quy

Lớp dư thừa

Lớp mật độ

Lớp chú ý

Method	Category	Video-type	Dataset	Architecture	Key idea
DeblurringNet et al. (2017)				DNet	Five neighboring blurry images are stacked and fed into a DNet to recover the center sharp image.
SPNet et al. (2017)				RNN-DNet	A recurrent architecture which includes a bi-recurrent temporal layer, facilitating the information propagation.
PCNN-Antenna and Temporal (2018)				DNet	Permutation invariant CNN which consists of several 1D filter, taking a sequence of blurry images as input.
DR-Net et al. (2018)				GAN	GAN-based video-defurring network, using a D-CNN to extract spatio-temporal information.
RefNet et al. (2018)	Non self-supervised	Monocular	Arranging	RefNet	Three consecutive blurry images are fed into the RefNetNet framework to generate sharp images, which are used to compute the optical flow and estimate a time-based flow, then reconstructing the sharp image.
DFNet et al. (2019)				RNN-DNet	Recurrent video-defurring network, where a hidden state is transferred from past frames to the current frame.
EDVR et al. (2019)				DNet	Pyramid, Causal and Deformable (PCD) module for frame alignment and a Temporal and Spatial Attention (TSA) module, facilitating reconstruction module to restore sharp images.
STNet et al. (2019)				DNet	STNet takes the current blurred frames, the preceding blurred and restored frames as input and generates a sharp version of the current frame.
CDNet-DP et al. (2020)				Unet-like	Convolutional deep video-defurring which first generates the denoising prior, then stacks both the blurry images and the prior into an UNet.

Hình: "Khái quát về phương pháp làm rõ video sâu

Kiến trúc mạng phổ biến nhất được sử dụng cho việc làm rõ ảnh

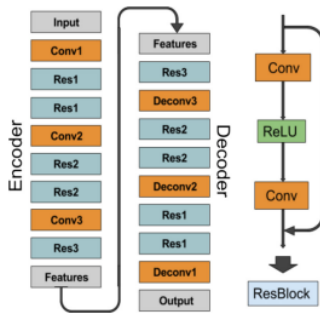
Deep Auto-Encoders (DAE)

Generative Adversarial Networks (GAN)

Cascaded Networks

Multi-Scale Networks

Reblurring Networks



Hình: Mạng làm rõ ảnh đơn sâu dựa trên kiến trúc Mã hóa Tự động Sâu (DAE) của^{9/21}

Các hàm mất mát

Các hàm mất mát

Hàm mất mát pixel

Hàm mất mát tri giác

Hàm mất mát đối kháng

Hàm mất mát tương đối

Hàm mất mát dòng chảy quang học

Các Bộ Dữ liệu Thử Nghiệm cho Việc Làm Rõ Ảnh

Bộ dữ liệu làm rõ ảnh

Bộ dữ liệu của Levin et al

Bộ dữ liệu của Sun et al

Bộ dữ liệu của Köhler et al

Bộ dữ liệu của Lai et al

Bộ dữ liệu của GoPro

Bộ dữ liệu HIDE

Bộ dữ liệu RealBlur

Bộ dữ liệu Blur-DVS

International Journal of Computer Vision

Table 4 Representative benchmark datasets for evaluating single image, video and domain-specific deblurring algorithms

Dataset	Synthetic	Real	Sharp Images	Blurred Images	Blur Model	Type	Train/Test Split
Levin et al. (2009)	×	✓	4	32	Uniform	Single image	Not divided
Sun and Hays (2012)	✓	×	80	640	Uniform	Single image	Not divided
Köhler et al. (2012)	✓	×	4	48	Non-uniform	Single image	Not divided
Lai et al. (2014)	✓	✓	108	300	Both	Single image	Not divided
GoPro Nish et al. (2017)	✓	✓	3,214	3,214	Non-uniform	Single image	2,103/1,111
HIDE Shih et al. (2019)	✓	×	8,422	8,422	Non-uniform	Single image	6,397/2,025
Blur-DVS Jiang et al. (2020)	✓	×	2,178	2,918	Non-uniform	Single image	1,782/396
Su et al. (2017)	✓	✓	6,708	11,925	Non-uniform	Video	5,708/1,000
REDS Nish et al. (2019)	✓	✓	30,000	30,000	Non-uniform	Video	24,000/3,000
Huadl et al. (2015)	✓	×	3M+35K	3M+35K	Non-uniform	Text	3M/35K
Shih et al. (2018)	✓	×	6,564	130M+16K	Uniform	Face	130M/16K
Zhou et al. (2019)	✓	×	20,637	20,637	Non-uniform	Stereo	17,319/3,318

Hình: Bộ dữ liệu thử nghiệm đại diện để đánh giá các thuật toán làm rõ ảnh đơn, video và trong các lĩnh vực cụ thể.

Bộ dữ liệu làm rõ video

DVD dataset: : Su et al. (2017) đã ghi lại các chuỗi video mờ bằng các máy ảnh trên các thiết bị khác nhau, bao gồm iPhone 5s, Nexus 4x và GoPro Hero 4. Họ đã lấy mẫu video sắc nét ở tốc độ khung hình 240 fps và lấy trung bình tám khung hình lân cận để tạo ra các video mờ tương ứng ở 30fps

Bộ dữ liệu REDS: Nah et al. (2019) đã ghi lại 300 đoạn video ở độ phân giải 1080×1920 với tốc độ khung hình 120 fps bằng máy ảnh GoPro Hero 6 Black và tăng tốc độ khung hình lên 1920 fps bằng cách sử dụng nội suy khung hình

Bộ dữ liệu cụ thể cho từng lĩnh vực

Bộ dữ liệu làm rõ văn bản: Hradiš et al. (2015) đã thu thập tài liệu văn bản từ Internet và giảm mẫu xuống độ phân giải 120–150 DPI

Bộ dữ liệu làm rõ khuôn mặt: Các bộ dữ liệu hình ảnh khuôn mặt mờ đã được xây dựng từ các bộ dữ liệu hình ảnh khuôn mặt hiện có

Bộ dữ liệu mờ stereo: Zhou et al. (2019) đã sử dụng một máy ảnh stereo ZED để ghi lại video 60 fps, được tăng lên 480 fps thông qua nội suy khung hình.

Đánh giá hiệu suất

Deblurring hình ảnh đơn: Các phương pháp đại diện được tổng kết trên bảng so sánh hiệu suất trên ba bộ dữ liệu deblurring hình ảnh đơn phổ biến, bao gồm bộ dữ liệu GoPro, bộ dữ liệu từ Köhler et al. và Shen et al.

- Deblurring hình ảnh UHD: Deblurring hình ảnh UHD là một nhiệm vụ khó khăn hơn, và các kiến trúc đa tỉ lệ đạt được hiệu suất tốt hơn.
- Deblurring lệch tiêu cự: Các phương pháp sử dụng mạng GAN thường đạt được hiệu suất tốt hơn cho deblurring lệch tiêu cự so với các kiến trúc đa tỉ lệ và đa miếng.

Deblurring video: Đánh giá hiệu suất: Các phương pháp deblurring video gần đây được so sánh trên hai bộ dữ liệu deblurring video phổ biến là bộ dữ liệu DVD và REDS. Các mô hình sử dụng mạng auto-encoder sâu là kiến trúc phổ biến nhất cho các phương pháp deblurring video.

So sánh với deblurring hình ảnh đơn: Các phương pháp deblurring hình ảnh đơn có thể được áp dụng cho video, nhưng chúng thường thua kém so với các phương pháp dựa trên video sử dụng thông tin thời gian.

Làm Rõ Ảnh Theo Miền Cụ Thể

Làm Rõ Ảnh Theo Miền Cụ Thể

Deblurring ảnh khuôn mặt: Các phương pháp deblurring ảnh khuôn mặt sử dụng cấu trúc chính của ảnh khuôn mặt và sự phát triển của deep learning. Sử dụng đặc điểm khuôn mặt cụ thể, các mạng học sâu đã chiếm ưu thế trong việc phục hồi ảnh khuôn mặt

Deblurring ảnh văn bản: Ảnh văn bản mờ có ảnh hưởng đến hiệu suất của các hệ thống nhận dạng ký tự quang học (OCR), ví dụ khi đọc tài liệu, hiển thị, hoặc biển chỉ dẫn đường phố. Các phương pháp deblurring ảnh chung không phù hợp cho ảnh văn bản

Deblurring ảnh stereo: Stereo vision được sử dụng rộng rãi để đạt được sự nhận biết độ sâu và hiểu biết về cảnh 3D. Tuy nhiên, khi lắp đặt một máy ảnh stereo trên một nền tảng di chuyển, sự rung lắc sẽ dẫn đến mờ, ảnh hưởng tiêu cực đến việc tính toán stereo tiếp theo.

Cơ hội và thách thức

Cơ hội và thách thức

Dữ liệu thực tế

Hàm mất mát

Các chỉ số đánh giá

Dữ liệu và mô hình

Các chi phí tính toán

Học không điều chỉnh



**HỌC VIỆN
CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

Posts and Telecommunications Institute of Technology

Cảm ơn thầy cô đã chú ý lắng nghe!

Sinh viên thực hiện: Nguyễn Thị Lan

Giáo viên hướng dẫn: ThS Đinh Xuân Trường

Mã sinh viên: B21DCCN818 - Lớp: D21CQCN09-B