

基于 SRCGAN 的超分辨率图像重建算法研究

毕玉萍 李劲华* 赵俊莉

(青岛大学数据科学与软件工程学院 山东 青岛 266000)

摘要 针对已有的方法存在超分辨率性能差和效率低等问题,提出一种新的基于生成式对抗网络的图像超分辨率算法。该方法融合条件生成式对抗网络(Conditional Generative Adversarial Nets, CGAN)的类标签作为生成式对抗网络(Generative Adversarial Networks, GAN)的输入,并训练 GAN 结构,从而实现图像的重建。在 MNIST 数据集和车牌数据集上进行对比实验,结果表明重建图像的准确率从原来的 68.22% 提升到 80.23%。

关键词 生成式对抗网络 超分辨率 条件生成对抗网络 图像处理

中图分类号 TP391.41

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2020.07.024

SUPER-RESOLUTION IMAGE RECONSTRUCTION ALGORITHM BASED ON SRCGAN

Bi Yuping Li Jinhua* Zhao Junli

(School of Data Science and Software Engineering, Qingdao University, Qingdao 266000, Shandong, China)

Abstract Aiming at the problems of poor super-resolution and low efficiency of the existing methods, this paper proposes a new image super-resolution algorithm based on the generative adversarial networks. It integrated the class labels of conditional generative adversarial nets (CGAN) as the input of generative adversarial networks (GAN), and GAN structure was trained to realize image reconstruction. The comparative experiments on MNIST datasets and license plate datasets show that the accuracy of reconstructing images is increased from 68.22% to 80.23%.

Keywords Generative adversarial networks Super-resolution Conditional generative adversarial nets Image processing

0 引言

图像超分辨率(Super Resolution, SR)重建一直以来是一个热门研究方向。它是将一幅或多幅低分辨率(Low Resolution, LR)图像作为输入,输出高分辨率(High Resolution, HR)图像的一种计算机视觉任务。该研究虽然具有较高的挑战性,但有许多应用价值,如医学图像处理^[1]、卫星成像^[2]、面部图像增强^[3-4]和压缩图像^[5]等领域。

目前,图像超分辨率算法已趋于成熟,主流的学习算法主要分为三种:基于插值^[6]的方法、基于重建^[7]的方法和基于学习^[8-10]的方法。基于插值的方法主要通过已有特征进行观测来获取像素点,通过这些

已有的像素点拟合出数学模型,从新建的数学模型中提取特征来重建图像。基于插值方法的算法存在计算量大、图像边缘和细节处理效果差的问题。基于重建的方法虽然可以重建较清晰的图像,但是存在计算量大、丢失高频细节的问题。基于深度学习的超分辨率图像重建技术主要有以下几种主要的方法:(1) 基于冗余字典建立低分辨率图片和高分辨率图片的特征块投影矩阵^[11-13]; (2) 通过建立数学模型实现低分辨率和高分辨率图像之间的一一对应关系^[11-12],主要方法有固定的领域回归、岭回归等模型。基于深度卷积神经网络的图像超分辨率模型^[14](Super Resolution Convolutional Neural Network, SRCNN)能够保持现有的重建效果,同时实现了较优的实验效果。该模型由三层卷积层组成,并分别负责特征提取、高分辨率特征表示

收稿日期:2019-06-12。国家自然科学基金项目(61702293);中国博士后科学基金项目(2017M622137);教育部虚拟现实应用工程研究中心基金项目(MEOBNUEVRA201601)。毕玉萍,硕士生,主研领域:深度学习,图像处理。李劲华,教授。赵俊莉,副教授。

的非线性映射以及高分辨率特征表示的非线性图像重建。SRCNN 模型是直接将低分辨率图像(LR)作为输入,输出高分辨率图像(HR),与已有的传统方法相比,处理比较简单。但该方法是一个浅层的网络结构,仍然存在着纹理模糊的问题。Ledig 等^[15]提出了基于生成对抗网络(GAN)的图像超分辨率方法(Super Resolution Convolutional Generative Adversarial Networks, SRGAN),该方法通过深层神经网络^[18-19]实现,在细节处理上取得更好的效果。

在 SRGAN 的基础上,本文提出一种改进的基于 GANs 的图像超分辨率算法,即具有类条件图像超分辨率重建框架(Super Resolution Conditional Generative Adversarial Nets, SRCGAN),SRCGAN 明确 GAN 类标签的条件下作为 GAN 的输入。本文通过为判别器添加这种辅助的自监督损失,可以实现更稳定的训练,使判别器表征对生成器输出质量的依赖性降低,从而提高重建效果。

1 算法设计

1.1 生成式对抗网络

近年来,GANs 作为生成模型的框架取得了相当大的成功。GAN 网络结构由两个模型组成:生成模型 G ,其作用是生成图像集,作为判别器的输入;判别模型 D ,其作用是辨别图像是来自真实图像还是生成器产生的伪图像。在 GAN 中,发生器和判别器被训练成一个双反馈网络结构,使生成的模型分布与真实分布之间的差异最小化。考虑输入噪声的先验分布 $p_z(z)$,为了了解生成器在数据 x 上的分布,由生成器执行从噪声分布到数据空间 $G(z; \theta_g)$ 的映射。GAN 定义判别器 $D(x; \theta_d)$,它输出的数值是来自生成器或真实数据集的概率。本文同时训练 G 和 D 以最小化对数 $\log(1 - D(G(z)))$ 和最小化 $\log(D(x|y))$ 。GANs 的标准公式如下:

$$\min_G \max_D V(G, D) = E_{x \sim P_{\text{data}}(x)} [\log D(x)] + E_{Z \sim P_Z(Z)} [\log(1 - D(G(z)))] \quad (1)$$

式中: P_{data} 是数据分布; P_Z 是通过对抗性 min-max 优化来学习的生成器分布。

1.2 条件生成对抗网络(CGAN)

Mirza 等^[18]在上述 GANs 的基础上提出了有条件的 GANs(CGANs),CGANs^[19-20]与 GAN 模型一样也是由生成器和判别器组成,但 GAN 无法控制数据的生成过程。为了解决此问题,研究者通过调整类标签上的模型,指导数据的生成过程。 y 被定义为类标签,此时

发生器和判别器被训练成一个双反馈回路的一部分,使生成的模型分布与真实分布之间的差异最小化。考虑输入噪声的先验分布 $p_z(z)$,为了了解生成器在数据 x 上的分布,由生成器执行从噪声分布到数据空间 $G(z|y; \theta_g)$ 的映射。GAN 定义判别器 $D(x|y; \theta_d)$,它输出的概率是判断来自生成器生成的伪数据还是真实数据。同时训练 G 和 D 在以 y 为条件的最小化对数 $\log(1 - D(G(z|y)))$ 和 D 最小化 $\log(D(x|y))$ 。有条件的 GANs 的标准公式如下:

$$\min_G \max_D V(G, D) = E_{x \sim P_{\text{data}}(x)} [\log D(x|y)] + E_{Z \sim P_Z(Z)} [\log(1 - D(G(z|y)))] \quad (2)$$

式中: P_{data} 是数据分布; P_Z 是通过对抗性 min-max 优化来学习的生成器分布。

1.3 SRCGAN

在 CGANs 和 SRGAN 的启发下,本文提出了一种具有类条件的图像超分辨率重建 GAN 框架(SRCGAN)。首先以比例因子对 HR 图像(IHR)进行降采样,若要创建 LR 输入图像(ILR),执行检测采样处理,随后给 GAN 提供 LR 图像和原始图像中的数字标签。与 SRGAN 相比,本文为 SRCGAN 框架加入了条件元素,这也是本文的改进之处。

生成器和判别器分别以 $D(x, y | \theta_d)$ 和 $G(x, y | \theta_g)$ 为条件,其中: x 表示 HR 图像(要么是真实图像,要么是生成器产生的图像)作为输入, y 是类标签。本文的对抗模式的目标是:

$$\min_G \max_D V(G, D) = E_{I^{HR} \sim P_{\text{data}}(I^{HR})} [\log D(I^{HR}, y)] + E_{I^{LR} \sim P_C(I^{LR})} [\log(1 - D(G(I^{LR}, y), y))] \quad (3)$$

式中: P_{data} 是数据分布; P_C 是通过对抗性 min-max 优化来学习的生成器分布。通过式(3)的检测采样处理,待优化的目标参数减少,同时加入类标签,让网络的优化训练更具针对性,同时也提高了低分辨率图像的重建效果。本文将条件 GANs 框架应用于 MNIST^[21]数据集,并表明它产生的图像更接近 MNIST 数字。改进的 SRCGAN 结构如图 1 所示,可知本文方法在下采样图片时加入了条件类,这些类标签是带有数据集特有属性的标签,再将处理过的图片输入到生成器,训练生成器生成的伪高分辨率的图片,生成器生成的伪高分辨率图片和真实的高分辨率图片输入到判别器中让判别器加以判别,当生成器生成的图片能够达到以假乱真的时候,模型训练完毕。本文的核心过程也是通过引入条件类解决在图像重建过程中随机处理的缺陷。通过条件类的引入,解决了重建过程中对于没有更正过

程的处理,从而进一步提升生成式对抗网络的重建效果。因此,无论在准确率还是适用范围,本文方法都有效地解决了目前重建效果差的问题。

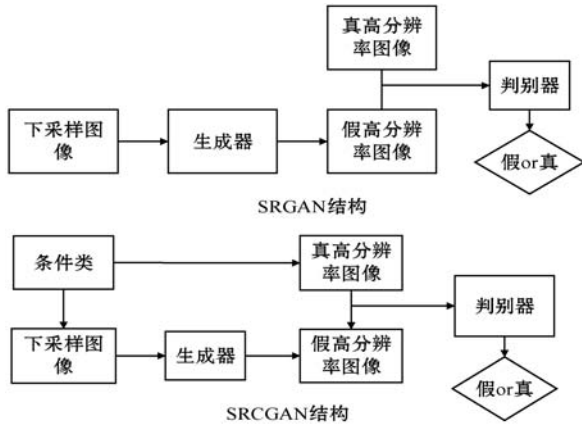


图1 SRCGAN 结构图与 SRGAN 结构图的对比

2 实验

2.1 数据集处理

本文使用 MNIST 数据集,共包含 70 000 幅图像,其中 60 000 幅用作训练集,10 000 幅用作测试集,数据集为手写体数字且服从均匀分布。数据集图像尺寸为 28×28 。首先用双三次插值将图像缩小到 7×7 (缩小 4 倍),作为实验的输入,同时输入数字标签作为条件变量,评估 SRCGAN 相对于传统 GAN (即没有任何条件变量)的性能。SRCGAN 训练的批次大小为 128,学习率为 0.001。该模型训练了 100 个 epochs (共迭代 46 875 次),利用 ReLU^[22] 作为激活函数,Adam^[23] 作为优化器建立模型。对 CNN 分类器进行训练,该模型训练了 70 个 epochs (共迭代 32 813 次),采用 ReLU 作为激活函数,Adam 作为优化器。两种网络结构都由 TensorFlow^[24] 库编写。

2.2 定量评估

由于选取的 MNIST 数据集不能通过人工验证实验效果,考虑到 CNN 分类器识别图片的高准确率,本文通过 CNN 分类器来验证实验结果。CNN 分类器的作用相当于一个评分器,通过这个评分器可以比较 SRGAN 与 SRCGAN 重建图片的准确率。训练 CNN 对 MNIST 数据集的数字进行分类,以评估来自 SRCGAN 的预测是否可以正确地被识别为实际数字。CNN 分类器在 MNIST 测试集上的准确率为 0.986 1。将 SRCGAN 和 SRGAN 生成的图像输入 CNN 分类器,判别分类精度分别为 0.802 3 和 0.682 2。这表明条件信息可以帮助生成更准确的 HR 图像。表 1 对比了 SRCGAN 和 GAN 生成的图像 MNIST 数字分类器的准确性。

表 1 SRCGAN 和 GAN 生成的图像在 MNIST 测试集上通过 CNN 分类器判别的准确度

模型	准确率/%
SRCGAN	80.23
GAN	68.22

2.3 定性结果

图 2 比较了 SRCGAN 和 GAN 的输出结果,图中第一列为输入的原始低分辨率图像,第二列为 GAN 重建的图像,第三列为 CGAN 重建的图像,第四列为原始高分辨率图像。由于 GAN 产生的大部分图像都非常模糊,因此很难分辨出图片具体代表什么数字;另一方面,用 SRCGAN 生成的数字图像更清晰,数字更容易识别。SRCGAN 的 HR 输出比缩小的图像有明显的改善。总之,基于 GAN 的比较分类器和可视化结果表明,加入条件项可以提高 HR 图像重建的质量。

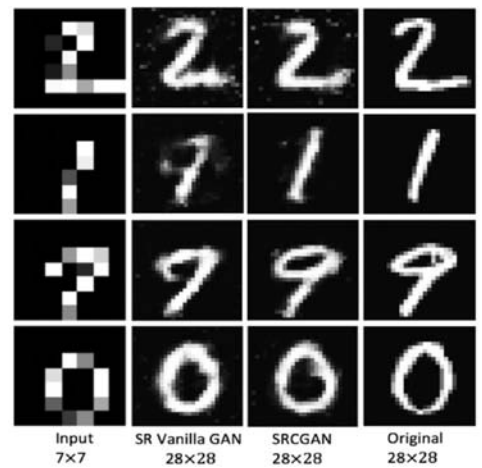


图 2 MNIST 分类器对 SRCGAN 输出和 GAN 输出的分类结果与 LR 输入图像和真实 HR 图像比较

2.4 车牌重建实验及其分析

为了进一步验证本文算法的应用价值,对车牌数据集进行了对比实验,设置并选取标准差为 10、均值为 0 的高斯分布来更新滤波器的初始值,并用随机梯度下降法得到目标函数的最优解。通过多次实验和调整参数,得到当放大因子为原图像的 3 倍时,图片重建效果表现最优。基于图片的输入,本文应用的主要方法为将 RGB 的颜色空间转换为 YCbCr 颜色空间中,通过对亮度 Y 通道的重建,其余则通过运用双三次插值法得到目标的最终尺寸。

本文选取了部分车牌图像进行处理,当放大因子为 3 的情况下将本文方法与双三次插值 Bicubic 算法和基于字典学习的 K-SVD 算法的重修效果作对比,结果如图 3 所示。其中第一列为原始高分辨率图像,第二列为 Bicubic 处理的图像,第三列为 K-SVD 处理后的图像,第四列为 SRCGAN 处理后的图像。

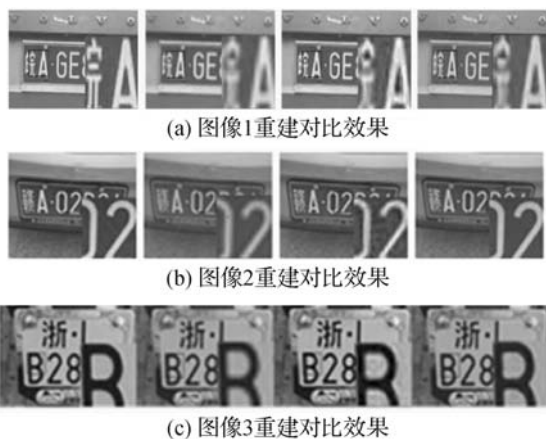


图3 放大因子为3的重建效果对比图

可以看出,本文方法优于其他两种算法,能够获得更加清晰的重建图片。在与 Bicubic 算法作对比时,本文算法得到的重建图像边缘更加清晰且更平滑,图片的整体图像视觉效果更好,且达到人眼的美学标准。而与 K-SVD 算法作对比时,本文算法可以完全克服 K-SVD 算法产生的振铃现象。车牌图像虽然经过本文算法得到的整体重建图片效果较好,但是在车牌图像重建后的中文文字部分还是不够清晰,有些细节信息不能很好地恢复和重建,与原始真实的高清车牌图像对比,重建的文字效果和真实的文字还存在一定差距。对整体车牌图像重建后也能高效清晰地重建车牌中的文字部分,是下一步研究工作中的重点和难点。

从客观角度分析,表2是对比不同算法后 PSNR 值的比较,当 PSNR 值越大,表示重建的效果越好。由表2可知,本文提出的重建算法得到的 PSNR 值明显高于其他两种算法,充分说明本文算法重建质量优于其他两种算法。Bicubic 算法和 K-SVD 算法虽然在一定程度上提高了车牌图像的质量,但是重修后的图像仍存在边缘模糊的短板和缺陷。对比实验表明:本文提出的 SRCGAN 模型重建后的车牌图像的边缘和纹理都更加清晰,且获得的图像的重修效果的 PSNR 值均高于 Bicubic 算法和 K-SVD 算法得到的 PSNR 值,整体效果更接近原始的高分辨率图像和真实图片。这表明了 SRCGAN 网络结构能够有效处理细节和边缘问题,显著提高图像的重建效果。

表2 不同算法对车牌重建后的 PSNR 值

图像	倍数	算法		
		Bicubic	K-SVD	SRCGAN
图像1	3	26.89	27.46	29.01
图像2	3	24.32	25.11	27.33
图像3	3	25.81	26.35	27.98

3 结 语

本文提出并实现了图像超分辨率的方法 SRCGAN。在 MNIST 数据集和车牌数据集上评估 SRCGAN 的性能后,成功地从 LR 输入中恢复了清晰的 HR 图像。模型输出与真实 MNIST 图像几乎一致,车牌的输出与真实图像也较为一致。SRCGAN 是明确 GAN 类标签的条件下作为 GAN 的输入。深层网络结构模型和 CNN 架构目前已能够表现出更优的性能,但是 SRCNN 只能在较大的图像尺度上进行训练,并且只能在特定尺寸的训练集上起作用。本文对 SRCGAN 进行改进,将类标签信息作为生成器的输入,可以指导生成器生成的图像更加接近数据集,从而提高图像的重建效率以及准确率。与无条件信息的 GAN 相比,本文方法让重建图片更精确且更接近真实图片。

参 考 文 献

- [1] Zikos M, Kaldoudi E, Orphanoudakis S C. DIPE: a distributed environment for medical image processing [J]. *Studies in Health Technology and Informatics*, 1997, 43:465-469.
- [2] Jacobsen K. High resolution satellite imaging systems-an overview[J]. *Photogrammetrie Fernerkundung Geoinformation*, 2005, 6(6): 487-496.
- [3] Terzopoulos D, Waters K. Analysis and synthesis of facial image sequences using physical and anatomical models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, 15(6): 569-579.
- [4] Wang G Q, Ou Z Y. Face recognition based on image enhancement and gabor features [C]//2006 6th World Congress on Intelligent Control and Automation. IEEE, 2006, 2: 9761-9764.
- [5] Tang J S, Peli E, Acton S. Image enhancement using a contrast measure in the compressed domain [J]. *IEEE Signal Processing Letters*, 2003, 10(10): 289-292.
- [6] Ji C T, He X H, Fu Y Q, etc. An edge directed interpolation algorithm based on regularization [J]. *Journal of Electronic Information*, 2014, 36(2): 293-297.
- [7] 孙京,袁强强,李冀玮,等.亮度——梯度联合约束的车牌图像超分辨率重建[J]. *中国图象图形学报*, 2018, 23(6): 802-813.
- [8] Schulter S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015, 3791-3799.
- [9] Salvador J, Pérez-Pellitero E. Naive bayes super-resolution forest [C]//2015 IEEE International Conference on Comput-

- er Vision, 2015: 325 – 333.
- [10] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution [C]//European Conference on Computer Vision. Springer, Cham, 2014: 184 – 199.
- [11] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections [J]. Journal of the Optical Society of America A, 1989, 6 (11): 1715 – 1726.
- [12] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014: 2672 – 2680.
- [13] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1646 – 1654.
- [14] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38 (2): 295 – 307.
- [15] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4681 – 4690.
- [16] Goodfellow I, Bengio Y, Courville A. Deep Learning [M]. Cambridge: The MIT Press, 2016: 26 – 29.
- [17] 曹仰杰, 贾丽丽, 陈永霞, 等. 生成式对抗网络及其计算机视觉应用研究综述 [J]. 中国图象图形学报, 2018, 23 (10): 1433 – 1449.
- [18] Mirza M, Osindero S. Conditional generative adversarial nets [EB]. arXiv preprint arXiv:1411.1784, 2014.
- [19] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1125 – 1134.
- [20] Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis [C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning, 2016: 1060 – 1069.
- [21] Yann L C, Cortes C, Burges C J C. The MNIST database of handwritten digits [OL]. <http://yann.lecun.com/exdb/mnist/>.
- [22] Xu B, Wang N Y, Chen T Q, et al. Empirical evaluation of rectified activations in convolutional network [EB]. arXiv preprint arXiv:1505.00853, 2015.
- [23] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB]. arXiv preprint arXiv:1412.6980, 2014.
- [24] Abadi M, Barham P, Chen J M, et al. Tensorflow: a system for large-scale machine learning [C]//12th USENIX Conference on Operating Systems Design and Implementation, 2016: 265 – 283.
- ~~~~~
- (上接第 106 页)
- [17] Li S, Gong D, Yuan Y. Face recognition using weber local descriptors [J]. Neurocomputing, 2013, 32 (9): 272 – 283.
- [18] Mehrsan J, Martin L. Online dominant and anomalous behavior detection in videos [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2013: 2609 – 2616.
- [19] Geng X, Yu C, Hu G. Unsupervised feature selection by kernel density estimation in wavelet-based spike sorting [J]. Biomedical Signal Processing and Control, 2012, 7 (2): 112 – 117.
- [20] Wright J, Yang A, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31 (2): 210 – 217.
- [21] Cai S, Zuo W, Zhang L. Support vector guided dictionary learning [C]//Computer Vision-ECCV 2014 – 13th European Conference, Zurich, Switzerland, September 6 – 12, 2014, Proceedings, Part IV, 2014: 183 – 202.
- [22] Andrade E, Fisher R. Modelling crowd scenes for event detection [C]//Proceedings of the 18th International Conference on Pattern Recognition (ICPR ' 06). 2006: 175 – 178.
- [23] Tal H, Yossi I, Orit K. Violent flows: Real-time detection of violent crowd behavior [C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2012.
- [24] Zhang T, Yang Z, Jia W, et al. A new method for violence detection in surveillance scenes [J]. Multimedia Tools and Applications, 2015, 2 (1): 1 – 23.
- [25] Dan X, Elisa R, Yan Y, et al. Learning deep representations of appearance and motion for anomalous event detection [C]//The British Machine Vision Conference (BMVC). 2015.
- [26] Dong Z, Qin J, Wang Y. Multi-stream deep networks for person to person violence detection in videos [C]//Chinese Conference on Pattern Recognition. Springer, Singapore, 2016: 517 – 531.
- [27] Meng Z, Yuan J, Li Z. Trajectory-pooled deep convolutional networks for violence detection in videos [C]//International Conference on Computer Vision Systems, 2017: 437 – 447.
- [28] Ding C, Fan S, Zhu M, et al. Violence detection in video by using 3D convolutional neural networks [C]//Advances in Visual Computing: 10th International Symposium, ISVC 2014, 2014: 551 – 558.