

遮蔽物除去による動物認識の精度向上

川中 康平^{1,a)} ラバリソン ミサマナナ フェリシア² 谷岡 広樹^{3,b)} 大平 健司³
佐野 雅彦³ 松浦 健二³ 上田 哲史³

概要：コンピュータで画像情報を用いて動物認識を行う際、対象となる動物の体の一部が檻などの遮蔽物に隠れ、動物認識の精度が低下する場合がある。一方、人間が視覚情報を用いて物体を認識する際、物体の一部が檻などの遮蔽物に隠れていても、遮蔽物を除去、補完し、物体を認識できることが知られている。そこで本研究では、コンピュータで画像から檻を除去することによる動物認識の精度向上を目指す。

キーワード：ノイズ除去、画像補完、動物認識

Improving Animal Recognition Accuracy using De-fencing

KOHEI KAWANAKA^{1,a)} RABARISON M. FELICIA² HIROKI TANIOKA^{3,b)} KENJI OHIRA³
MASAHIKO SANO³ KENJI MATSUURA³ TETSUSHI UETA³

Abstract: When a computer recognizes an animal on an image, an accuracy of animal recognition decreases if a part of the animal's body is covered by fences. On the other hand, when a human recognizes a object on the image, it is known to be possible to remove and supplement the fences, and recognize the object, even if a part of the object is covered by fences. In this research, we aim to improve the accuracy of animal recognition using de-fencing on computer.

Keywords: Noice reduction, Image complement, Animal recognition

1. 序論

画像認識の精度は年々向上している。2010年から、画像認識や画像分類の精度を定量的に測り、競う ILSVRC (ImageNet Large Scale Visual Recognition Competition) が始まった。このコンテストにおいて、2012年に Hinton らがディープラーニングによる画像認識 [1] を行い、画像認識の精度は大幅に向上した。2011年までのコンテスト

では画像処理技術を用いた画像認識が一般的であり、画像認識のエラー率は26%前後であった。ところが、Hinton らのチームのエラー率は16%で画像処理技術を用いた画像認識に圧勝した(図1)。更に、2016年にはコンピュータによる画像の認識率が人間の認識率を上回った [2]。

一方、日常生活において、撮影した写真に映る物体の一部が遮蔽物によって隠れている場面は多々ある。このような場合、コンピュータによる画像情報を用いた画像認識では、認識率が低下してしまう可能性がある。人間はこのような問題に直面した場合も、脳で遮蔽物を除去・補完し物体を認識できることが知られている [3]。そこで本研究では、コンピュータでも人間の脳と同様に遮蔽物を除去・補完することで、物体の認識率の向上を目指す。具体的には、檻によって体の一部が隠れている動物の写真を対象に、檻を除去することによる動物認識の精度向上を目指す。

¹ 徳島大学 大学院先端技術科学教育部
Graduate School of Advanced Technology and Science,
Tokushima University

² 徳島大学 工学部知能情報工学科
Department of Information Science and Intelligent Systems,
Tokushima University

³ 徳島大学 情報センター
Center for Administration of Information Technology,
Tokushima University

a) 1201.kouhei@gmail.com

b) tanioka.hiroki@tokushima-u.ac.jp

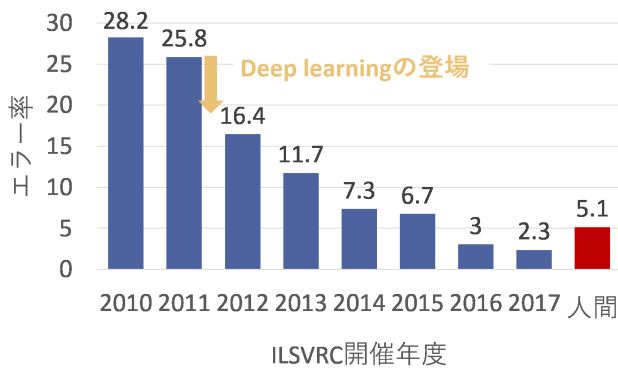


図 1 ILSVRC におけるエラー率
Fig. 1 Error rate of ILSVRC

2. 関連研究

2.1 画像認識技術

ILSVRC では、2012 年に Hinton らが AlexNet による画像認識で精度を大幅に向上して以降、ディープラーニングを用いた画像認識が盛んにわれている。中でも有名な 3 つの画像認識の手法を以下に示す。

1 つ目は、2014 年の ILSVRC で 2 位になった Oxford 大学の VGG チームが開発した VGG [4] である。AlexNet と同じ考えで層を深くしたもので、畳み込み層とプーリング層を連結した単純な構造であるため、現在でもよく使われている。2 つ目は、2014 年の ILSVRC で優勝した Google による GoogLeNet [5] である。こちらは、それまでのネットワークは縦に深い構造であったのに対して、インセプションモデルと呼ばれる横に広い構造を提案している。異なるサイズのフィルターの畳み込み層を横に並べて、出力層でそれらを 1 つに結合している。3 つ目は、2015 年の ILSVRC で優勝した、Kaming He らによる ResNet [6] である。ディープラーニングでは、ネットワークを深くしすぎると性能が落ちるといった問題があったが、Kaming He らは、この問題に対してスキップ構造を用いて解決した。スキップ構造とは、ある層への入力をその前の層をスキップし、逆伝播の際の勾配の消失や発散を低減する構造である。前年優勝の GoogLeNet が 22 層であったのに対し、ResNet では 152 層の深さのネットワークであった。

2.1.1 檻の認識

島村らは、檻が直線であるため輝度差の勾配方向の双対性を示すこと、また人工物はフラクタル性を持っているという特徴から檻などの細長い人工遮蔽物の検出を行っている [7]。しかし、この方法では檻だけでなくメガネなどの細長い物体を検出してしまう。これに対して、Liu らは檻のパターンを認識し、檻を検出する [8]。そのため、メガネなどは検出されることがない。その他、檻の認識に CNN-SVM を用いる提案 [9] などもある。

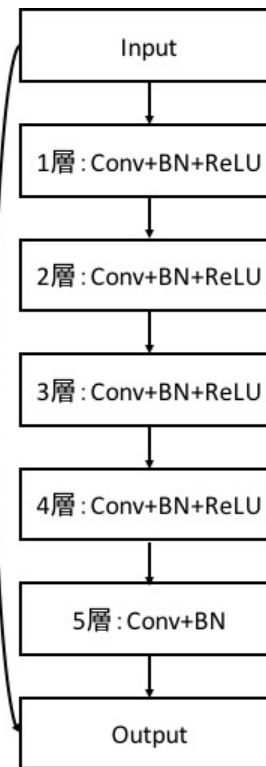


図 2 Win5-RB のモデル
Fig. 2 Win5-RB model

2.1.2 檻部分の補完

画像処理ライブラリ OpenCV では、画像補完のために inpaint 関数が含まれる。inpaint 関数には Telea のアルゴリズム [10] と Bertalmio らのアルゴリズム [11] が利用できる。Telea は Fast Marching Method を用いて欠損箇所の周囲から内側に向かって徐々に補完する。Bertalmio らは流体力学に基づいた考え方で、偏微分方程式を利用したアルゴリズムで欠損箇所を補完する。飯塚らは、畳み込みニューラルネットワークを用いて画像補完を行う手法を提案している [12]。

3. 提案手法

3.1 檻の除去システム

このため、従来の手法では、まず檻を認識し、次に補完するという手順で檻などの遮蔽物除去 (de-fencing) をおこなうことが多かったが、本研究では、ディープラーニングを用いたノイズ除去の手法の 1 つである Win5-RB [13] を用いて、End-to-End の学習を行うことで、檻の除去と認識した檻部分の補完を同時におこなう。Win5-RB は 図 2 で示すように Wide Inference Network (Win) を 5 層構成の ResNet (R) と BatchNormalization (B) を使用して、ノイズ除去を行う目的で開発された手法である。檻をノイズとみなして学習させることで、檻を除去する推論システム (以下、遮蔽物除去システム) を実現する。図 3 に提案手法のフローチャートを示す。

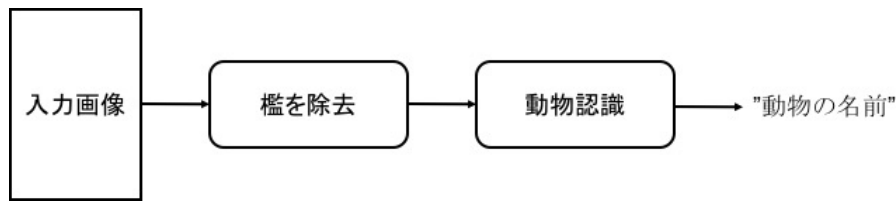


図 3 提案手法のフローチャート
Fig. 3 Flow chart of proposed method

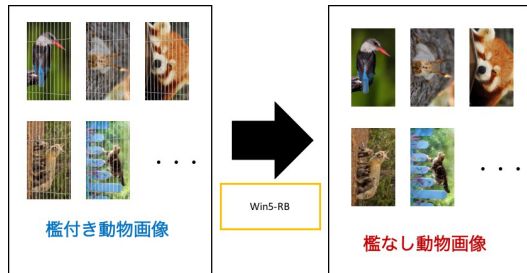


図 4 遮蔽物除去システム的作用
Fig. 4 Role of de-fencing system.



図 7 3枚の檻あり画像
Fig. 7 Three fenced images.

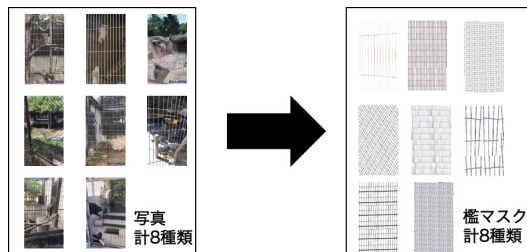


図 5 檻マスクの作り方
Fig. 5 How to make a fence mask.

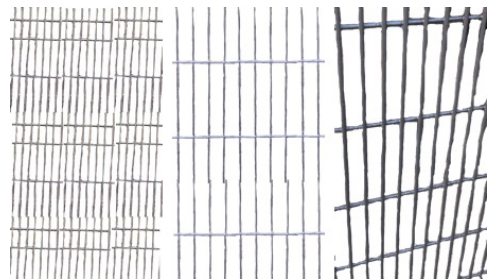


図 8 3種類の檻マスク画像
Fig. 8 Three types of fence mask images.

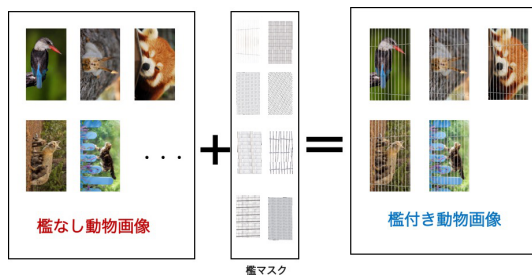


図 6 擬似的な檻あり画像の作り方
Fig. 6 How to make pseudo fenced images.

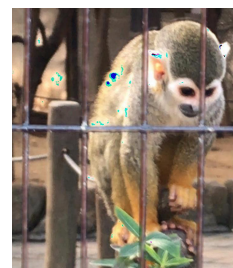


図 9 推論結果のノイズ
Fig. 9 Noise caused using Win5-RB.

3.2 学習データの作成

遮蔽物除去システムの学習では、図 4 のように、対象となる動物の前景に檻がある画像（以下、檻あり画像）を入力し、動物の前景に檻がない画像（以下、檻なし画像）が出力されるように学習する。しかしながら、檻なし画像と全く同一の条件で撮影された檻あり画像は存在しないため、図 5 のように、檻あり画像から手動で檻部分のみを抜き出し、図 6 のように、檻なし画像に檻マスクを重ねることで、擬似的な檻あり画像を作成し、入力データとする。

3.3 推論モデルの生成

遮蔽物除去システムの学習には、スマートフォンによる写真撮影時に一般的な画像サイズである $1,920 \times 1,080$ px の画像を用いる。このとき、メモリ消費量に問題で、Win5-RB で画像サイズ $1,920 \times 1,080$ px のカラー画像を学習させること困難であったため、 320×280 px の画像サイズで、32枚に分割したものを使用する [14]。画像分割した擬似的な檻あり画像と、これに対応する画像分割した檻なし画像のセット 3,999 組を学習に使い、40 epochs 実行する。

表 1 擬似的な檻あり画像からの檻除去の効果 (RMSE, SSIM)

Table 1 Effectiveness of de-fencing pseudo fenced images (RMSE, SSIM).

画像	RMSE	SSIM
擬似的な檻あり画像	10.2411	0.1999
擬似的な檻除去画像	6.8391	0.6996

Win5-RB による推論の結果は、本来 0 から 1 の範囲 [0, 1] の値を取る。しかし、図 9 のようなノイズが入ることがあったため、本実験では、0 未満の値は全て 0 とし、1 より大きい値は全て 1 とした。

4. 動物認識の実験

4.1 檻の除去精度の評価

まず、遮蔽物除去システムによる檻の除去精度を検証する。檻なし画像と同一の条件で撮影した檻あり画像は存在しない。そのため、檻あり画像から手動で檻部分のみを抜き出した檻のマスクを、檻なし画像に重ね、擬似的な檻あり画像を作成し、実験を行う。

4.1.1 評価用データセット

評価のために、30 枚の檻なし画像 (図 10) と、3 枚の檻あり画像 (図 7) をデータセットとして準備する。檻なし画像には、VGG-16 が学習画像として使っている大規模画像セットである ImageNet に含まれるニホンザル、リスザル、ワオキツネザルの 3 種類の動物の画像から、それぞれ 10 枚ずつを選ぶ。擬似的な檻あり画像の作成のためには、とくしま動物園で撮影されたニホンザル、リスザル、ワオキツネザルの 3 枚の檻あり画像から、檻部分を切り出して、檻マスク (図 8) を 3 枚の作成する。これらの檻マスクを 30 枚の檻なし画像にランダムに重ねることで、擬似的な檻あり画像 (図 11) 30 枚を作成する。

4.1.2 除去精度の評価指標

評価指標には、RMSE (Root Mean Squared Error) と SSIM (Structural SIMilarity) の 2 つの指標を用いて評価をおこなう。MSE (Mean Squared Error) は、実際の値と予測値の 2 乗を平均したもので、平均二乗誤差と呼ばれている。式 (1) で表す RMSE は MSE (Mean Squared Error) の平方根のことであり、値が 0 に近いほど 2 枚の画像は類似しているといえる。

$$\text{RMSE}(x, y) = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (x_i - y_i)^2} \quad (1)$$

ここで、 n は画素数、 x_i は檻なし画像 x の画素、 y_i は遮蔽物除去システムの出力画像 (檻除去画像) y の画素を表す。次に、式 (2) で表す SSIM は MSE より人間の主観的判断に近いとされる画像評価の指標である。輝度やコントラスト、構造を軸にして周囲のピクセル平均、分散、共分散を取ることでピクセル単位ではなく、周囲のピクセルとの相関を含んだ評価をする。

表 2 檻なし画像と擬似的な檻あり画像の比較 (NDCG)

Table 2 Comparison of un-fenced images with pseudo fenced images (NDCG).

画像	NDCG
檻なし画像	0.5207
擬似的な檻あり画像	0.3517

表 3 擬似的な檻除去の比較 (NDCG)

Table 3 Comparison of pseudo fenced images (NDCG).

画像	NDCG
擬似的な檻あり画像	0.3559
擬似的な檻除去画像	0.3670

表 4 実際の檻除去画像の比較 (NDCG)

Table 4 Comparison for real fenced images (NDCG).

画像	NDCG
実際の檻あり画像	0.2648
実際の檻除去画像	0.2958

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

ここで、 μ_x は檻なし画像 x の平均、 μ_y は檻なし画像 y の平均、 σ_x^2 は檻なし画像 x の分散、 σ_y^2 は檻なし画像 y の分散、 σ_{xy} は x と y の共分散、 $L = 2^{32} - 1$ 、 $k_1 = 0.01$ 、 $k_2 = 0.02$ 、 $c_1 = (k_1 L)^2$ 、 $c_2 = (k_2 L)^2$ とする。

4.1.3 除去精度の評価結果

檻なし画像 30 枚と擬似的な檻あり画像 30 枚の RMSE、SSIM の平均値、檻なし画像 30 枚と擬似的な檻除去画像 30 枚の RMSE、SSIM の平均値を比較する。表 1 より、檻なし画像と擬似的な檻除去画像の RMSE は、檻なし画像と擬似的な檻あり画像の RMSE と比較し、3.402 低くなっている。また、檻なし画像と擬似的な檻除去画像の SSIM は、檻なし画像と擬似的な檻あり画像の RMSE と比較し、0.4997 高くなっている。以上の結果から、遮蔽物除去システムを用いて檻を除去することで、擬似的な檻画像を檻なし画像に近づけることができた。

4.2 動物認識の精度比較

次に、認識対象の動物の前景に檻がある場合と、檻がない場合を比較して、動物認識の精度が低下することを確認するために、檻なし画像と檻あり画像に対する動物認識の精度を比較する。評価用データセットには、4.1.1 章で用いたものを使用する。

4.2.1 動物認識の精度指標

動物認識には keras [15] に実装されている VGG-16 の学習済みモデルを用い、認識率の上位 5 位 ($k = 5$) までの物体名を表示する。精度評価には式 (3) で表す NDCG (Normalized Discounted Cumulative Gain) を用いる。



図 10 檻なし画像の例

Fig. 10 Examples of un-fenced images



図 11 擬似的な檻あり画像の例

Fig. 11 Examples of pseudo fenced images

$$NDCG \equiv DCG = \sum_{i=1}^k \frac{rel_i}{\log_2(i+1)}. \quad (3)$$

このとき、 i は認識結果内の正解ラベルの順位、 rel_i は認識率である。NDCG は DCG [16] を正規化した値であり、0 から 1 の範囲 $[0, 1]$ の値をとる。本実験においては、画像 1 枚あたり正解ラベルは 1 つであるため、NDCG と DCG は同値となり、NDCG (DCG) の値が 1 に近いほど正確に動物を認識できていることを表している。

4.2.2 動物認識精度の比較結果

表 2 は、檻なし画像 30 枚と擬似的な檻あり画像 30 枚の NDCG の平均値を比較したものである。擬似的な檻あり画像の NDCG は、檻なし画像の NDCG と比較して 0.1690 低く、動物の認識精度が低下していることが分かる。

4.3 檻除去画像の動物認識の精度評価

擬似的な檻あり画像と、擬似的な檻あり画像の檻除去画像の動物認識精度を比較する。評価用データセットには、4.1.1 章で用いたものを使用する。また、評価指標は NDCG とする。表 3 より、檻を除去することで擬似的な檻あり画像よりも NDCG の値が 0.0111 向上し、認識率が向上していることが分かる。

4.4 実際の檻あり画像との精度評価

12 は実際の檻あり画像である。この画像から遮蔽物除去システムを用いて檻を除去した画像（以下、実際の檻除去画像）を作成する。作成した実際の檻除去画像は、図 13

である。評価指標には NDCG を用いる。表 4 より、実際の檻除去画像と実際の檻あり画像の動物認識精度を比較すると、檻を除去することで NDCG の値が 0.0310 向上し、認識率が高くなっていることが分かる。

5. 結論

人間の物体認識の方法に習い、檻が重なった動物の画像に対して、遮蔽物除去システムを用いて檻を除去し、動物認識の実験を行なった結果、檻を除去しない場合と比較して動物認識の精度が向上することが分かった。一方、本実験に用いた Win5-RB による障害物除去システムでは、訓練データに含まれない檻のパターンについては、適切に除去できない。学習時により多くの檻のパターンを学習させることで、現在よりも檻の除去精度の向上が期待できる。さらに、実際のアプリケーションとして実装し、動物園などでフィジビリティテストを実施したい。

謝辞 本研究に用いるデータセットは、とくしま動物園北島建設の森で撮影したものを使用している。ここに記して深謝する。

参考文献

- [1] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems 25* (Pereira, F., Burges, C. J. C., Bottou, L. and Weinberger, K. Q., eds.), Curran Associates, Inc., pp. 1097–1105 (online), available from <http://papers.nips.cc/paper/4824->



図 12 実際の檻あり画像の例

Fig. 12 Examples of real fenced images



図 13 実際の檻除去画像の例

Fig. 13 Examples of de-fenced images

- imagenet-classification-with-deep-convolutional-neural-networks.pdf) (2012).
- [2] He, K., Zhang, X., Ren, S. and Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034 (online), DOI: 10.1109/ICCV.2015.123 (2015).
 - [3] Ban, H., Yamamoto, H., Hanakawa, T., Urayama, S.-i., Aso, T., Fukuyama, H. and Ejima, Y.: Topographic Representation of an Occluded Object and the Effects of Spatiotemporal Context in Human Early Visual Areas, *Journal of Neuroscience*, Vol. 33, No. 43, pp. 16992–17007 (online), DOI: 10.1523/JNEUROSCI.1455-12.2013 (2013).
 - [4] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *International Conference on Learning Representations* (2015).
 - [5] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826 (online), DOI: 10.1109/CVPR.2016.308 (2016).
 - [6] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pp. 770–778 (online), DOI: 10.1109/CVPR.2016.90 (2016).
 - [7] 島村 潤, 数藤恭子, 志水信哉, 谷口行信: 勾配方向の双対性とフラクタル次元を用いた細長い人工遮蔽物の検出, *映像情報メディア学会誌*, Vol. 67, No. 7, pp. J219–J228 (オンライン), DOI: 10.3169/itej.67.J219 (2013).
 - [8] Liu, Y., Belkina, T., Hays, J. and Lublinerman, R.: Image de-fencing, *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, (online), DOI: 10.1109/CVPR.2008.4587493 (2008).
 - [9] Jonna, S., Nakka, K. K. and Sahay, R. R.: Deep Learning Based Fence Segmentation and Removal from an Image Using a Video Sequence, *Computer Vision – ECCV 2016 Workshops* (Hua, G. and Jégou, H., eds.), Cham, Springer International Publishing, pp. 836–851 (2016).
 - [10] Telea, A.: An Image Inpainting Technique Based on the Fast Marching Method, *Journal of Graphics Tools*, Vol. 9, No. 1, pp. 23–34 (online), DOI: 10.1080/10867651.2004.10487596 (2004).
 - [11] Bertalmio, M., Bertozzi, A. L. and Sapiro, G.: Navier-stokes, fluid dynamics, and image and video inpainting, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pp. I–I (online), DOI: 10.1109/CVPR.2001.990497 (2001).
 - [12] Iizuka, S., Simo-Serra, E. and Ishikawa, H.: Globally and Locally Consistent Image Completion, *ACM Trans. Graph.*, Vol. 36, No. 4, pp. 107:1–107:14 (online), DOI: 10.1145/3072959.3073659 (2017).
 - [13] Liu, P. and Fang, R.: Wide Inference Network for Image Denoising, *CoRR*, Vol. abs/1707.05414 (2017).
 - [14] Wu, S., Zhang, M., Chen, G. and Chen, K.: A New Approach to Compute CNNs for Extremely Large Images, *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17*, New York, NY, USA, ACM, pp. 39–48 (online), DOI: 10.1145/3132847.3132872 (2017).
 - [15] Chollet, F. et al.: Keras, <https://keras.io> (2015). Accessed on Feb. 8, 2019.
 - [16] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N. and Hullender, G.: Learning to Rank Using Gradient Descent, *Proceedings of the 22Nd International Conference on Machine Learning, ICML '05*, New York, NY, USA, ACM, pp. 89–96 (online), DOI: 10.1145/1102351.1102363 (2005).