Deep Manifold Alignment for Mid-grain Sketch based Image Retrieval

Tu Bui[1], Leonardo Ribeiro[2], Moacir Ponti[2], and John Collomosse[1]

[1] University of Surrey — Guildford, Surrey GU2 7XH, UK
{t.bui,j.collomosse}@surrey.ac.uk
[2] Universidade de São Paulo — São Carlos/SP 13566-590, Brazil
{ leonardo.sampaio.ribeiro,ponti}@usp.br

Abstract. We present an algorithm for visually searching image collections using free-hand sketched queries. Prior sketch based image retrieval (SBIR) algorithms adopt either a category-level or fine-grain (instance-level) definition of cross-domain similarity — returning images that match the sketched object class (category-level SBIR), or a specific instance of that object (fine-grain SBIR). In this paper we take the middle-ground; proposing an SBIR algorithm that returns images sharing both the object category and key visual characteristics of the sketched query without assuming photo-approximate sketches from the user. We describe a deeply learned cross-domain embedding in which 'mid-grain' sketch-image similarity may be measured, reporting on the efficacy of unsupervised and semi-supervised manifold alignment techniques to encourage better intra-category (mid-grain) discrimination within that embedding. We propose a new mid-grain sketch-image dataset (MidGrain65c) and demonstrate not only mid-grain discrimination, but also improved category-level discrimination using our approach.

Keywords: SBIR · Manifold alignment · Visual Search.

1 Introduction

Free-hand sketch offers an intuitive and convenient query modality for visual search when a photographic sample of the desired content is unavailable. Yet, matching sketches and photographs is challenging; sketches are salient abstractions frequently drawn from canonical viewpoints, caricaturing objects, and introducing non-linear deformations [4, 21]. Recently deep neural networks, in particular multi-branch (triplet) networks, have proven effective in learning a mapping across these two domains for sketch based image retrieval (SBIR). Such approaches typically fall into either of two camps according to granularity at which matching is performed: 1) category (object-level) SBIR in which a sketched query of a given object (e. g. a cat) should return images containing that object (e. g. cats) [1–3]; 2) fine-grain (instance-level) search in which a detailed sketch of a specific object (e. g. a shoe) should return only that specific shoe [21, 16]. Whilst both bodies of work have made significant advances in cross-domain (sketch-photo) matching, arguably neither provides a model for practical SBIR. Category-level matching is analogous to sketched object classification, suggesting