Image Processing and Scientific Misconduct

William R. Oliver
Armed Forces Institute of Pathology

In a recent microscopy discussion group, a correspondent questioned the use of a contrast enhancement method applied to a microscopy image for publication. What kinds of image processing, folk asked, are "acceptable" in a general sense? What kinds of image processing should be noted in the text or caption and what kind of images should be archived?

It is an interesting set of questions, and I face them wearing three hats. I am a forensic pathologist who performs forensic image interpretation in the investigation of homicide and assault, an anatomic pathologist with an interest in confocal microscopy, and a computer scientist with training in image processing and computer vision. Image processing is part and parcel of my everyday work. I use tools such as contrast enhancement, deblurring, and photogrammetry for image interpretation. I build tools for visualization in my confocal work. I collaborate in the design of data acquisition devices for the evaluation of crime scenes and bodily injury.

The questions of reporting and archiving can be approached as both methodological and ethical problems. Unfortunately, many folk seem to be awfully quick to turn methodological disagreements into accusations of ethical wrongdoing. They are very different things to me.

The practical response to problems of documentation and archiving can be complicated, even for the best-intentioned. For publication, different journals have different explicit instructions to authors, or ignore the problem altogether. For instance, the journal *Science* requires that "any materials and methods necessary to verify the conclusions of the experiments be made available to other investigators under appropriate conditions" and that "archival data sets (such as sequence and structural data) should be deposited with the appropriate data bank"[1]. The *American Journal of Clinical Pathology* requires that "established techniques may be referenced; however new or modified methods should be described in sufficient detail to allow duplication of the study by an independent observer."[2]

The requirements for presentation in court are a little different. The biggest determinant is whether or not the expert opinion is based on the processed image or whether the processed image is used merely to illustrate an opinion. If the processed image is an inherent part of the expert opinion, then there are fairly rigorous rules for admission into court. Until a few years ago, expert opinion and supporting data had to pass the "Frye test" which stated that any technique must be generally accepted by the scientific community as shown by publication in peer-reviewed journals and such. More recently. the Supreme Court held that the Frye test was too "austere"[3]. It noted that scientific thought was not static. Instead, Justice Blackmun noted that "scientists do not assert they know what is immutably true – they are committed to searching for new, temporary theories to explain, as best they can, phenomena."[4] Using this reasoning the Supreme Court gave individual judges much more freedom in allowing evidence and placed the responsibility of determining credibility on the jury. The Daubert decision applied only to federal courts. Some State courts have adopted the Daubert rule, and some have retained Frye. In contrast, if the image is used as an illustration and is not part of the process which led to the expert opinion, rules for admission are much more liberal. Challenges tend to center on whether or not the illustration appeals more to the jury's emotions than reason, or whether or not the illustration is germane.

But, really, all the instructions to authors and all the Supreme Court decisions in the world don't tell us what is ethical. They merely tell us what is procedurally required in a specific instance. How can a scientist, academic or forensic, make decisions in some consistent manner? My answer is that, in spite of the wishes of the more Pharisaically-inclined amongst us, there is likely no single answer for all situations. Instead, one must use a little professional judgement instead of looking in an ethical cookbook.

Most simply, I try not to mislead people and I try to provide enough information and archive enough data to make my work reproducible. As simple as this rule is, however, it can be a little complicated in practice. When is it misleading to process an image? To what depth should I report the kind of image processing that I do on an image? Is it enough to state an algorithm? Is it enough to state an algorithm and the parameters? Is it necessary to publish the pseudocode or explicit C++ code with every article?

Clearly it is not "wrong" to process an image. All images are necessarily processed, whether chemically or digitally. Some, such as CT or MRI images are fundamentally artificial and involve a profound amount of processing in their very creation. It is no more "wrong" to optimize an image using digital tools than it is to get the best possible results in a darkroom. The fundamental question, to me, is whether or not the manipulation changed the results upon which the conclusion is based. If, for instance, the conclusion is based on the absence or presence of a feature, then any enhancement which makes the feature more visible is fair game as long as the processing doesn't create or destroy the features in question. On the other hand, if a conclusion is based on the relative value of two features, then any manipulation which changes relative values would obviously be misleading.

I hold myself to a similar rule for the level of reporting when providing an image for publication or evidence. I don't report basic calibration, color balance optimization or things like contrast stretches if they are used to make features more easily viewed but are not an inherent part of the conclusion, any more than I report the graphics algorithms used to create bar graphs and scatter plots. If I perform some image manipulation which I don't think will be intuitively obvious, but which still has no affect on the conclusion, I will mention that it was done and perhaps reference the manipulation in passing, but will not go into detail about the particulars of the algorithm. If the underlying conclusion is based on quantification of image features or if the image manipulation affects the result in some quantitative way, as with a densitometric measurement on Feulgen stained cells, then every step from illumination correction to noise reduction to how the measurements are done must be reported.

But again, even in that last case, it's not as simple as it sounds. My goal is to describe image processing algorithms with enough detail so that my colleagues can reproduce the results, and not so much that I bore them to tears. But it's not always obvious how much a colleague needs - and thus what is "adequate."

If I say "I performed a contrast-limited adaptive histogram equalization (CLAHE) on the image," that is, as far as I'm concerned, an accurate description of what I did. Most folk who do image processing know what an CLAHE does. However, I'll bet there are as many different implementations of AHE as there are folk who have written AHE code. To be "accurate" do I need to say that I used 12 horizontal divisions and 10 vertical divisions? Do I need to add that I used a contrast limiting factor of "4"? Do I need to provide pseudocode to tell folk what "4" means, or can I assume that anybody who knows what a CLAHE is can figure that out? Do I need to tell folk whether I used whole numbers only (integers) or allowed real numbers (floating point numbers)? Do I need to tell them that my program used numbers with 64 bits of precision as opposed to 32 or 16 or 8 or 128 bits?

The bottom line is that there comes a point where you stop and say that, hey, this is enough, and folk ought to be able to get it from here. One can be wrong in good faith. I once had a colleague who had trouble reproducing a result of mine. It turned out that one of the steps involved subtracting one image from another. My software represented image values as floating point numbers, and could handle negative values. My colleague assumed integer values between 0 and 255. When I subtracted a pixel value of 100 from one of 50, I got a result of -50. My colleague got a result of 205 because of the way his software handled negative numbers. My colleague gave me a call and we figured it out. We didn't accuse each other of scientific misconduct. Was I unethical because I didn't consider the possibility that folk would not be able to handle negative image values and thus didn't explicitly report that my subtractions could result in negative numbers? Of course not.

Similarly, is it enough to archive the original data and the final product, or

should one archive every intermediate image along the way? If you answer that one should archive every intermediate step along the way, then what constitutes the "smallest reportable intermediate step?" If I incorporate 15 steps into one program and push one button, does that make it one step? If I write 15 small programs and pipe intermediate images between them to accomplish the same thing, does that make it 15 steps? Does pushing one button on a commercial product constitute one step? What about if the commercial product is a development environment like AVS or Khoros? If I explore the results of parameterization of an algorithm by turning a dial, do I have to archive an image every fraction of a dial turn? Once again, one has to just use some common sense and decide what one thinks is enough for a reasonable colleague. I archive my original data, my final data, a description of the manipulations and parameterizations. I save intermediate images only when I think they will save me time if I decide to reanalyze the data.

The most insidious conflict of assumptions I have seen among my colleagues has been when an illustration is presented to show one point, but the reader inferred a point not made by the illustration at all. In the case I remember the best, an author performed a histogram equalization on an electrophoresis result to show the presence of a band. The equalization did not change whether or not the band was present, and thus did not misrepresent the point made in the paper. The reader. however, inferred an incorrect quantitative relationship between the band in question and the other bands and felt misled. Some folk have called the use of histogram equalization in cases like this unethical. Since I believe that this label implied the motivation to mislead, I disagree. I'd be willing to bet that, had the author foreseen that folk would make an incorrect inference, he or she would have made a note about the manipulation.

For the sake of clarity and consideration, it is best to err on the side of documentation, as has always been the case. However, you simply cannot foresee all possible questions about configuration, background, and assumptions. And those assumptions can kill you. When two people work from different assumptions, that doesn't mean that one of them must be unethical. Calling people unethical is highly loaded. It implies that people were consciously misleading and forces folk into battling camps. It might be better to approach disagreements in a less accusatory fashion. Everybody makes inaccurate assumptions when communicating with others on occasion. Everybody makes mistakes. And we can all correct our mistakes. Once somebody is labeled as "unethical," that label becomes quite a millstone.

Of course, real scientific fraud and misconduct does happen. However, most folk simply make honest mistakes or engage in honest miscommunication. People should be quick to question results and challenge methods. That's fun. But they should be slow to turn every such challenge into an accusation of unethical behavior or scientific misconduct. I don't look forward to the day when my colleagues start calling me "unethical" because I didn't correctly second-guess their assumptions. I'd much rather argue about it over a beer. ■

1) Instruction for Authors. Science, http ://www.sciencemag.orglmi scicon-info. shtml.
2) Information for Authors, American J Clin Path, 108(1):119,1997.
3) Frye v. U.S. (Frye), 293 F. 1013, 1014 (D.C. Cir. 1923)
4) Daubert v. Merrell Dow Pharmaceuticals (Daubert), 61 U.S.L.W 4805 (U.S. June 29,1993)

The opinions expressed in this article are those of the author only and do not necessarily represent those of the Armed Forces Institute of Pathology, the United States Army, the government of the United States.